

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

Om nucleic - nucleic search, using sw model

Run on: September 8, 2002, 04:53:07 ; Search time 217.98 Seconds  
(without alignments)  
8664.125 Million cell updates/sec

```
Title: US-09-896-791B-2
Perfect score: 1100
Sequence: 1 gaattcgccagcagggccat.....aaaaaaacatcgcgcgc 1100
```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

```
Searched:      1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

N.genseq/ncb/2802: \*

- 1: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1980.DAT: \*
- 2: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1981.DAT: \*
- 3: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1982.DAT: \*
- 4: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1983.DAT: \*
- 5: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1984.DAT: \*
- 6: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1985.DAT: \*
- 7: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1986.DAT: \*
- 8: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1987.DAT: \*
- 9: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1988.DAT: \*
- 10: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1989.DAT: \*
- 11: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1990.DAT: \*
- 12: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1991.DAT: \*
- 13: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1992.DAT: \*
- 14: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1993.DAT: \*
- 15: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1994.DAT: \*
- 16: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1995.DAT: \*
- 17: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1996.DAT: \*
- 18: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1997.DAT: \*
- 19: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1998.DAT: \*
- 20: /SIDS1/gcgdata/genseq/genseqn-emb1/NA1999.DAT: \*
- 21: /SIDS1/gcgdata/genseq/genseqn-emb1/NA2000.DAT: \*
- 22: /SIDS1/gcgdata/genseq/genseqn-emb1/NA2001A.DAT: \*
- 23: /SIDS1/gcgdata/genseq/genseqn-emb1/NA2001B.DAT: \*
- 24: /SIDS1/gcgdata/genseq/genseqn-emb1/NA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	556.6	50.6	2207	20	AAK56986	Mouse transcript14
2	471.6	42.9	2082	22	AAH14706	Human CDNA sequenc
3	447.4	40.7	2595	22	AAH17091	Human CDNA sequenc
4	418.8	38.1	711	22	AAH07034	Human CDNA clone (
5	392.8	35.7	756	22	AAH03708	Human CDNA clone (
6	253.4	23.0	1708	22	AAH15203	Human CDNA sequenc
7	253.2	23.0	873	22	AAH03186	Human CDNA clone (
8	211.6	19.2	733	22	AAH03810	Human CDNA clone (
9	211.6	19.2	1841	22	AAH15364	Human CDNA sequenc

10	174.6	15.9	2730	20	AA588981	Human transcribed
11	173	15.7	1607	21	AA294052	Human endothelial
12	173	15.7	2816	19	AAV00641	CDNA encoding human
13	173	15.7	2818	21	AA294053	DNA encoding human
14	173	15.7	33169	22	ABA14358	Human nervous syst
15	172	15.6	2776	22	AA514145	Human EPAS1 DNA us
16	156.6	14.2	3031	19	AAV00642	CDNA encoding huma
17	138.2	12.6	3746	24	AB199710	Mouse ischemic co
18	134.6	12.2	2528	22	AA514154	Human HIF-1 alpha
19	134.6	12.2	3678	19	AA563210	Nucleic acid sequ
20	134.6	12.2	3736	18	AAAT45937	Human hypoxia indu
21	134.6	12.2	3736	21	AA2939537	DNA encoding a wil
22	134.6	12.2	3927	20	AA565690	Lung small cell ca
23	134.6	12.2	3933	20	AA565890	Human transcribed
24	134.6	12.2	10355	22	AA514156	PSMART CMV-HIF DNA
25	102.2	9.3	4162	23	AA585058	DNA encoding novel
26	97.4	8.9	2487	23	AA585055	DNA encoding novel
27	92	8.4	139	21	AAAC27335	Human secreted pre
28	89.6	8.1	2078	19	AAV41256	Human neuronal PAS
29	83.2	7.6	4565	21	AA233295	Drosophila melanog
30	83.2	7.6	4571	23	AB111895	Drosophila melanog
31	83.2	7.6	12776	23	AB111894	Drosophila melanog
32	79	7.2	1475	21	AAE18197	Lung cancer associ
33	72.6	6.6	3614	18	AAE191883	Murine SIM (singl
34	72.2	6.6	238	18	AAE191880	Partial human SIM
35	72	6.5	2082	19	AAV41257	Mouse neuronal PAS
36	69.2	6.3	549	21	AAAC93991	Cat files hindgut a
37	68.4	6.2	861	21	AAA40542	Human fetal brain
38	67.6	6.1	561	22	AAH09235	Human cDNA clone (
39	66.4	6.0	27884	22	AAE37781	Human Immune/haem
40	65.6	6.0	411	22	AAE193716	Human polyuncleot
41	65.6	6.0	416	22	AAE188617	Human lymphocyte
42	65.6	6.0	1432	24	AB134259	Human immune syst
43	65.2	5.9	449	22	AAE16186	Human breast canc
44	65.2	5.9	1835	21	AAE18015	Lung cancer associ
45	65.2	5.9	2058	22	AA526005	Human cDNA encodin

## ALIGNMENTS

XX	RESULT	1	
XX	AAx58986		
ID	AAx58986	standard; cDNA; 2207	BP.
XX	AAx58986;		
AC			
XX			
DT	23-AUG-1999	(first entry)	
XX			
DE	Mouse transcription regulator MOP7	cDNA.	
XX			
KM	MOP7; member of the PAS superfamily; bHLH-PAS; mouse;		
XX	transcription regulator; hypoxia inducible factor 3 alpha; ss		
OS	Mus musculus.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	84..2072	
FT		/*tag= a	
XX			
PN	W09928464-A2.		
XX			
PD	10-JUN-1999.		
XX			
PF	27-NOV-1998;	98MO-US25314.	
XX			
PR	28-NOV-1997;	97US-0068863.	
XX			
PA	(WISC ) WISCONSIN ALUMNI RES FOUND.		
XX			
PI	Bradfield CA, Gu YZ, Hogenesch JB;		
XX			
DR	WPI; 1999-371120/31.		

DR P-PSDB; AAY06295.

XX Developmental signal transduction associated proteins  
XX  
XX Claim 6; Page 100-101; 106pp; English.

PS This is the nucleotide sequence of a cDNA encoding MOP7 (see  
XX AAY06295), a novel member of the PAS superfamily, where PAS stands  
XX for PER/ARNT/SIM domains. The cDNA was identified in a search of  
CC murine ESTs designed to identify basic-helix-loop-helix-PAS  
CC (bHLH-PAS) proteins, and by RACE amplification of mouse lung cDNA.  
CC MOP7 has been characterised as hypoxia inducible factor 3 alpha  
CC (HIF 3 alpha). Its expression profile is distinct from that of  
CC HIF 1 alpha (see AAY06289), HIF 2 alpha (see AAY06290), MOP3 (see  
CC AAY06291), Ah receptor and Ah receptor nuclear translocator (ARNT),  
CC suggesting a different functional role. MOP7 probably regulates  
CC the same genes as HIF 1 alpha and 2 alpha, as evidenced by its  
CC dimerisation with the same partners (ARNT, MOP3) and recognition  
CC of the same core response element. MOP7 may have a functional  
CC role associated with response to low oxygen in the tissues in  
CC which it is expressed. The invention provides novel MOPs 2-9  
CC nucleic acids (see AX58981-88) and proteins (see AAY06289-97).  
CC These are useful in a variety of research, diagnostic and  
CC therapeutic applications. Several of the MOPs are alpha-class  
CC hypoxia-inducible factors. Others are involved in circadian signal  
CC transduction.

XX Sequence 2207 BP; 471 A; 742 C; 597 G; 397 T; 0 other;

Query Match 50.6%; Score 556.6; DB 20; Length 2207;  
Best Local Similarity 88.1%; Pred. No. 2.8e-105;  
Matches 658; Conservative 0; Mismatches 19; Indels 70; Gaps 2;

QY 26 tggggtcgtcagcgtcgaagtgcgaacacccgagctgcggaaggaagtcgcggagcgg 85  
DB 85 tggactcgggacacagagctcgaacacccgagctgcggaaggaagtcgcggagcgg 144  
QY 86 ccgc 145  
DB 145 ccgc 204  
QY 146 ttgc 205  
DB 205 ttgc 264  
QY 206 accctgc 251  
DB 265 accctgc 324  
QY 252 agccactgc 311  
DB 325 agccactgc 384  
QY 312 agggagacatgctcctcctcgtcggaatgtaagaacacactgggctcagtcgtga 371  
DB 385 agggagacatgctcctcctcgtcggaatgtaagaacacactgggctcagtcgtga 440  
QY 372 cccctctcctcctcctcctcgtataataccacactcctgtgtaaatctctctgtgaag 431  
DB 441 -----ctggagct 448  
QY 432 catgtgacacagtatcttgaattatcatccctctgtgacaaagaaacttaaacgc 491  
DB 449 catgtgacacagtatcttgaattatcatccctctgtgacaaagaaacttaaacgc 508  
QY 492 cctgaccccccagcgaacctcgtcaaaagaagaagctggaagcccccacagagccactt 551  
DB 509 cctgaccccccagcgaacctcgtcaaaagaagaagctggaagcccccacagagccactt 568  
QY 552 ttcctctggaatgaaagagacgctcaccagcagaagggcgcaagctcaactcaaacgcgc 611  
DB 569 ttcctctggaatgaaagagacgctcaccagcagaagggcgcaagctcaactcaaacgcgc 628

QY 612 cactcgtgaaggtcgtcgaactgctcgaacatatgagggcctacaagccccctgcacagac 671  
DB 629 cactcgtgaaggtcgtcgaactgctcgaacatatgagggcctacaagccccctgcacagac 688  
QY 672 ttcctctgaggaagcctcgtcgcagcctccctcgtcaatgctggtgtatctatctgta 731  
DB 689 ttcctctgaggaagcctcgtcgcagcctccctcgtcaatgctggtgtatctatctgta 748  
QY 732 agccatccccccagctccctcttcacaga 758  
DB 749 agccatccccccagctccctctgta 775

#### RESULT 2

AAH14706  
ID AAH14706 standard; cDNA, 2082 BP.  
XX  
XX AAH14706;  
AC  
XX 26-JUN-2001 (first entry)  
DT  
XX  
DE Human cDNA sequence SEQ ID NO:12421.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 12421; 2537bp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination  
CC of the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 2082 BP; 416 A; 684 C; 607 G; 375 T; 0 other;

Query Match 42.9%; Score 471.6; DB 22; Length 2082;  
 Best Local Similarity 79.9%; Pred. No. 8.4e-88;  
 Matches 612; Conservative 0; Mismatches 84; Indels 70; Gaps 2;

QY 7 ggcacagagggcactgagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 66  
 DB 1 gactgcgagacatgagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 60  
 QY 67 gagaatcgcgagagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 126  
 DB 61 gaaaatcccggttggggtcagcggttgaagttcgaacaacagctgcggaag 120  
 QY 127 ctggcgacactctgacgttggggtcagcggttgaagttcgaacaacagctgcggaag 186  
 DB 121 ctggcgacactctgacgttggggtcagcggttgaagttcgaacaacagctgcggaag 180  
 QY 187 atgcgctacaactgacgttggggtcagcggttgaagttcgaacaacagctgcggaag 232  
 DB 181 atgcgctacaactgacgttggggtcagcggttgaagttcgaacaacagctgcggaag 240  
 QY 233 caggtggaagaaaggagagagcactgagcggttgaagttcgaacaacagctgcggaag 292  
 DB 241 caggtggaagaaaggagagagcactgagcggttgaagttcgaacaacagctgcggaag 300  
 QY 293 gtaatgttactcagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 352  
 DB 301 gtaatgttactcagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 360  
 QY 353 ctgggctcagtcagtcagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 412  
 DB 361 ctgggctcagtcagtcagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 420  
 QY 413 ccaattctctctgagtcacttgaacaacagctgcggttgaagttcgaacaacagctgcggaag 472  
 DB 376 -----ctggagctcacttgaacaacagctgcggttgaagttcgaacaacagctgcggaag 424  
 QY 473 aagaggaacttaagagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 532  
 DB 425 aagaggaacttaagagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 484  
 QY 533 ccccaacagagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 592  
 DB 485 ccccaacagagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 544  
 QY 593 cgcctaacctcaaaagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 652  
 DB 545 cgcctaacctcaaaagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 604  
 QY 653 acaagccctctgacaggttggggtcagcggttgaagttcgaacaacagctgcggaag 712  
 DB 605 acaagccctctgacaggttggggtcagcggttgaagttcgaacaacagctgcggaag 664  
 QY 713 gctctgttactatctggaagcagctgcggttgaagttcgaacaacagctgcggaag 758  
 DB 665 gctctgttactatctggaagcagctgcggttgaagttcgaacaacagctgcggaag 710

RESULT 3

AAH17091  
 ID AAH17091 standard; CDNA; 2595 BP.

XX AAH17091;

XX 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:16414.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 KW Homo sapiens.  
 XX EP1074617-A2.  
 PN 07-FEB-2001.  
 PD 28-JUL-2000; 2000EP-0116126.  
 XX 29-JUL-1999; 993P-0248036.  
 PR 27-AUG-1999; 993P-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX (HELI-) HELIX RES INSP.  
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI; 2001-318749/34.  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 PS Claim 8; SEQ ID 16414; 2537pp + CD ROM; English.  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

SO Sequence 2595 BP; 497 A; 897 C; 703 G; 498 T; 0 other;

Query Match 40.7%; Score 447.4; DB 22; Length 2595;  
 Best Local Similarity 80.0%; Pred. No. 8.4e-83;  
 Matches 583; Conservative 0; Mismatches 76; Indels 70; Gaps 2;

QY 44 gtcgacacacgagctgaggaagagagtcgagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 103  
 DB 83 gtcgacacacgagctgaggaagagagtcgagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 142  
 QY 104 aggaagacgagctgctacacagctgagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 163  
 DB 143 aggaagacgagctgctacacagctgagcggttggggtcagcggttgaagttcgaacaacagctgcggaag 202  
 QY 164 cgcactgagcaagagctcactacatgcgctcacaatcagctacgctgcggttgaagttcgaacaacagctgcggaag 223  
 DB 203 cgcactgagcaagagctcactacatgcgctcacaatcagctacgctgcggttgaagttcgaacaacagctgcggaag 262

QY	224	ctctgagc-----agcaagctgaaaaaagggagagccacctgacgcctgct	269
Db	263	ctctgagcgcagaggagctggaacaaagctggagcaaaaggaggagaaacacctggaatgcctgct	322
QY	270	acctgaagagccctcgagagggttctctcatgtgtactcacgcgcagaggagagacatgctaac	329
Db	322	acctgaagagccctcgagagggtcttctcatgtgtactcacgcgcagaggagagacatgctaac	382
QY	330	tgctggaaaaatgtcaagaagcaccttgggcctcagtcagtcagtcactgttctcctccct	389
Db	383	tgctggagaaatgtcaagaacacaccttggtcctcagtcag-----	420
QY	390	gatacaaacccacacccctcgtaaccaattctctcttgagagctaatgtgagacagatctt	449
Db	421	-----cctgagagctcaattgtgacacagatctt	446
QY	450	tgatttaccatccctctgtgacccaagagaaacttcaagaagcccttgaccctcccaagccgaa	509
Db	447	tgatttaccatccacccctctgtgacccaagagagcttcaagaagcccttgaccctcccaagcagac	506
QY	510	ctgttcaaaagaagaagcttggaagccccaacagagcgcacattctccttgagaaatgaag	569
Db	507	ccgtctcagagaggaagtgtagagagccccaacagagcggctgtcttctccttgagaaag	566
QY	570	caagctccacagagagagggcgacgcttcaacccctcaaaagcggccacccctggaaagtgcctga	629
Db	567	tacgctccacacagcgcggcgacccctcaaacctccaagcggccacccctggaaagtgcctga	626
QY	630	ctgctcagagacataatgagggcctacaaagcccccctgcacagaacttcccttgccggagagcc	689
Db	627	ctgctcagagacataatgagggcctacaaagcccccctgcagaaacttcccttgagtgagagcc	686
QY	690	tcgcttcgaagcctccctcgaatgctgtgtctatctgtgaaagcaatccccacagctccc	749
Db	687	tgactcaagagcccccgctcagtgctgtgtctatctgtgaaagcaatccccacagag	746
QY	750	cttccacga	758
Db	747	cagccttga	755
RESULT 4			
AAH07034			
ID	AAH07034 standard; cDNA; 711 BP.		
AC	AAH07034;		
XX	26-JUN-2001 (first entry)		
DT			
XX	Human cDNA clone (5'-primer) SEQ ID NO:3869.		
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss		
XX	Homo sapiens.		
OS			
XX	EPI074617-A2.		
PN			
XX	07-FEB-2001.		
PD			
XX	28-JUL-2000; 2000EP-0116126.		
PF			
XX	29-JUL-1999; 99JP-0248036.		
PR	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
PR	02-MAY-2000; 2000JP-0183767.		
PR	09-JUN-2000; 2000JP-0241899.		
XX			
PA	(HELI-) HELIX RES INST.		
XX			
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX	WPI; 2001-318749/34.		

xx Primer sets for synthesizing polynucleotides, particularly the 5602  
 pn full-length cDNAs defined in the specification, and for the detection  
 pn and/or diagnosis of the abnormality of the proteins encoded by the  
 pt full-length cDNAs -  
 ps Claim 1; SEQ ID 3869; 2537P + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-*or* primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialized methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AA982446 to AA985893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 711 BP; 149 A; 239 C; 209 G; 111 T; 3 other;

Query Match	38.1%	Score 418.8;	DB 22;	Length 711;
Best Local Similarity	76.7%	Pred. NO. 4.7e-77;		
Matches 581; Conservative	0;	Mismatches 84;	Indels 73;	Gaps 4;

[illegible]

```

QY 533 ccccaacagagcgccactttctcctggaatgaagacagctacc-accagagagcg 591
    ||||| ||||| ||| ||| ||||| ||||| ||||| ||| |||||
Db 485 cccccagagcggtgtctctctctgcatgaagtaagctaccacagcgcgcg 544
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 592 accgtcaactcaaaagcgccactggaaggtgtctgacgcgtcagacataagggc 651
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 545 accctcaactcaagcgcgccactggaaggtgtctgacgcgtcgtacataagggc 604
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 652 tacaagccccctgacagacttccctgcgagagcctgcctcgag--cctccctgc 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 605 tacaagccactgtgcaactcttcancatcggaaccctgactcaaancccccgcgca 664
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 710 aatgcctggtgctatct 727
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 665 atgcctggtgctatct 682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
AAH03708 standard; cDNA; 756 BP.
XX
AC AAH03708;
XX
DE 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:543.
XX
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
KM Homo sapiens.
XX
OS
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 543; 2537bp + CD ROM; English.
XX
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the

```

```

CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 756 BP; 148 A; 263 C; 218 G; 123 T; 4 other:
XX
Query Match 35.7%; Score 392.8; DB 22; Length 756;
Best Local Similarity 79.3%; Pred No. 1, ie-71;
Matches 571; Conservative 0; Mismatches 74; Indels 75; Gaps 6;

QY 44 gttcgaacaccgagctgaggaaggaagtcgagcgcgcccgagcgcgagcc 103
    ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83 gttcgaacaccgagctgaggaaggaagtcgagcgcgcccgagcgcgagcc 142
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 104 aggaagacgaggtgtgttaccagctgagcgacactctgccccttgccgcggtacag 163
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 143 aggaagacgaggtgtgttaccagctgagcgacactctgccccttgccgcggtacag 202
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 164 cgcacctggaacagcgctccatcatgctgcctcacactacagctgcatgacgcgc 223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 203 cgcacctggaacagcgctccatcatgctgcctcacactacagctgcatgacgcgc 262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 224 tctgcgc-----agcaggtggaanaagggagagacactgtagcgcgtgt 269
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 263 tctgcgcgcgagggagtggaaccaggttggaagagggaggaacactgtagcgcgtgt 322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 270 acctgaagcgccctggaaggttctgcatgtgtaccacgcgcggagggagatgtgtacc 329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 323 acctgaagcgccctggaaggttctgcatgtgtaccacgcgcggagggagatgtgtacc 382
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 330 tgtcgaanaatgtcagcaagcactggtgcctcagtgtagcctctgtcctcctcct 389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 383 tgtcgaanaatgtcagcaagcactggtgcctcagtgtagcctctgtcctcctcct 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 390 gatacataaccacactcctgtgtaccaattctctctggaagctcatgtgacaagatcct 449
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 -----ctggaagctcatgtgacaagatcct 446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 450 tgatttatcatcctcctgtgaccagaaggaacttcaagaagccttgacccccaagcgaa 509
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 447 tgatttatcatcctcctgtgaccagaaggaacttcaagaagccttgacccccaagcgaa 506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 510 cctgtcaagaagaagctgtgaagccccaacagagcgccacttccctgctgaatgaagag 569
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 507 cctgtcaagaagaagctgtgaagccccaacagagcgccacttccctgctgaatgaagag 566
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 570 cagctcaaccagcagagcgacgctcaacctcaaaagcgagcactgtgaagtgtgtga 629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 567 tacgtcaaccagcagcgcgccaccccaacctcaaaagcgagcactgtgaagtgtgtga 625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 630 ctgtctggaacatatgagggcgctaccagcc--ccttgacagacttccctgcccggagggc 688
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 626 ctgtctggaacatatgagggcgctaccagccacacttcccaacttcccaactgtgagggc 685
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 689 ctgcctcgagcctccc--ctgcaatgctgtgtgttatc-tgtgaagccatcccccagc 745
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 686 ctgacccaacaaaccccgnttgcaatgtcgtgtctatcttgtgaagccatcccccac 745
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
AAH15203 standard; cDNA; 1708 BP.
XX
ID AAH15203
XX
AC AAH15203;
XX
DE 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:13302.

```



CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 873 BP, 175 A; 305 C; 248 G; 142 T; 3 other;

Query Match 23.0%; Score 253.2; DB 22; Length 873;

Best Local Similarity 86.6%; Pred. No. 5.7e-43;

Matches 279; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 424 ctgagctcatggaacacagatcttgatttaccacccctgtagcaagaagaactt 483  
DB 281 ctgagctcatggaacacagatcttgatttaccacccctgtagcaagaagaactt 340  
QY 484 caagagccctgaaccccaagccgaacctgtcaagaagaagctggaagcccaagag 543  
DB 341 cagagagccctgaaccccaagccgaacctgtcaagaagaagctggaagcccaagag 400  
QY 544 cgcacatttccctcgaaatgaagagcaagctcaccaagaagggcgacgctcaaccc 603  
DB 401 cgtgtcttctctctgctgactgaagtaagctcaccaagcgggcgacccctcaaccc 460  
QY 604 aagcgcgcaacctggaaggtgctgactgctcaagacatatgagggctacaagccct 663  
DB 461 aagcgcgcaacctggaaggtgctgactgctgacatatgagggctacaagccact 520  
QY 664 gacagagacttccctcgcgagagccctgctccgagccctccctggaagctgtgtctt 723  
DB 521 ggcgcagacttctcagctctggaagccctgactcagagccccgcgtcagtgctgtctc 580  
QY 724 atctgtaagccatccccagc 745  
DB 581 atctgtaagccatccccagc 602

RESULT 8

AAH03810  
ID AAH03810 standard; cDNA; 733 BP.

XX AAH03810;

AC 26-JUN-2001 (first entry)

DE Human cDNA clone (5'-primer) SEQ ID NO:645.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

XX EP1074617-A2.

PN 07-FEB-2001.

PD 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX

PS Claim 1; SEQ ID 645; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination  
CC of the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 733 BP; 143 A; 264 C; 204 G; 119 T; 3 other;

Query Match 19.2%; Score 211.6; DB 22; Length 733;

Best Local Similarity 83.1%; Pred. No. 2e-34;

Matches 241; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 472 caagagaaacttcaagagccctgaccccaagcgcaacctgtcaagaagaagctgaa 531  
DB 272 caagagaaacttcaagagccctgaccccaagcgcaacctgtccaggaaggtgag 331  
QY 532 gccccaagagagcgcaaccttccctcgcaatgaagaagcctcagcaggaagcgcc 591  
DB 332 gccccaagagagcggtgtcttccctcgcaatgaagaagcctcagcaggaagcgcc 391  
QY 592 acgctcaacctcaagaagcgccacacttgaaggtgctgactgctcagacatatgagggc 651  
DB 392 accctcaacctcaagaagcgccacacttgaaggtgctgactgctgacatatgagggc 451  
QY 652 tacaagccctcgcaagaagcttccctcgcggaagcgccctcgctcgaagctccctcgaa 711  
DB 452 tacaagccctcgcggaagcttccctcgcggaagcgccctcgctcgaagctccctcgaa 511  
QY 712 tgctgtgtctatctgtgaagccatccccagctccctccaagatg 761  
DB 512 tgctgtgtctatctgtgaagccatccccagctccctccaagatg 561

RESULT 9

AAH18364  
ID AAH18364 standard; cDNA; 1841 BP.

XX AAH18364;

AC 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:18402.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

XX EP1074617-A2.

PN 07-FEB-2001.

PD 28-JUL-2000; 2000EP-0116126.

XX

PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Oka T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishi S, Sugiyama T, Makamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
PS  
PS Claim 8: SEQ ID 18402; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX  
SQ Sequence 1841 BP; 404 A; 514 C; 519 G; 404 T; 0 other;  
Query Match 19.2%; Score 211.6; DB 22; Length 1841;  
Best Local Similarity 83.1%; Pred. NO. 2.4e-34;  
Matches 241; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 472 caaaggaactcaagaacgacctgaccgccagcgaacctgtcaagaagaagctggaa 531  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 272 caaaggaagctcagcagcgcctgagcccccagcagacccgttcaggaaggaagtgag 331  
532 gcccaacagagcgccaccttccctgcgaatgaagaagacagctcaccagcagagcgcc 591  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 332 gccccaagagcggtgcttctccttgcgcatgaagagtaagctcaccagcgcgcgagc 391  
QY 592 acggtcaacctcaaaagcgccacctggaaggtgtcgtcagctcgaagacatagggcc 651  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 392 accctcaacctcaaaagcgccacctggaaggtgtcgtcagctcgtgacatatagggcc 451  
QY 652 tacaagcccttgacacagacttccctcgtcgagagcctcgtccgcagacccctccctgaa 711  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 452 tacaagcaccactcgacagacttctccagcttggaagccttgagctcagaccccgctgag 511  
QY 712 tgctgtgtgtctatctgtgaagcattcccccctccctccacagatgg 761  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 512 tgctgtgtgtctatctgtgaagcattcccccctccctccacagatgg 761  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
RESULT 10  
AAK58981

ID AAK58981 standard; cDNA; 2730 BP.  
XX  
XX AAK58981;  
AC  
XX  
XX 23-AUG-1999 (first entry)  
DT  
XX  
XX Human transcription regulator MOP2 cDNA.  
DE  
XX  
XX MOP2; member of the PAS superfamily; bHLH-PAS; human;  
KM transcription regulator; hypoxia inducible factor 2 alpha; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS 33..2645  
FT CDS /\*tag- a  
FT  
XX  
XX MN0928464-A2.  
PN  
XX  
XX 10-JUN-1999.  
PD  
XX  
XX 27-NOV-1998; 98WO-US25314.  
PF  
XX  
XX 28-NOV-1997; 97US-0066863.  
PR  
XX  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
PA  
XX  
XX Bradfield CA, Gu YZ, Hogenesch JB;  
PI  
XX  
XX WPI; 1999-371120/31.  
DR  
XX  
XX P-PSDB; AAY06290.  
DR  
XX  
XX Developmental signal transduction associated proteins  
PT  
XX  
XX  
PS  
XX  
XX Claim 6; Page 94-95; 106pp; English.  
CC  
CC This is the nucleotide sequence of a cDNA encoding MOP2 (see  
CC AA06290), a novel member of the PAS superfamily, where PAS strands  
CC for PBR/ARNT/STW domains. The cDNA was identified in an iterative  
CC search of human ESTs designed to identify basic-helix-loop-helix-PAS  
CC (bHLH-PAS) proteins that interact with either the Ah receptor (AHR)  
CC or the Ah receptor nuclear translocator (ARNT). To obtain extended  
CC open reading frames for each EST, an anchored-PCR strategy was used  
CC to amplify additional flanking sequences from a commercial HepG2  
CC library. MOP2 appears to be related structurally and functionally  
CC to MOP1 (see AA06289). It interacts with ARNT, but not AHR, and  
CC drives transcription in its ARNT-dimerised form. Unlike MOP1, it  
CC does not appear to interact significantly with HSP90. MOP2 is  
CC induced by low oxygen and may be involved in hypoxia responses in  
CC difference cells and tissues. MOP2 is sometimes referred to as  
CC hypoxia inducible factor 2 alpha. The invention provides novel  
CC MOPs 2-9 nucleic acids (see AAK58981-88) and proteins (see AA06289-97).  
CC These are useful in a variety of research, diagnostic and  
CC therapeutic applications. Several of the MOPs are alpha-class  
CC hypoxia-inducible factors. Others are involved in circadian signal  
CC transduction.  
XX  
XX  
SQ Sequence 2730 BP; 659 A; 839 C; 727 G; 505 T; 0 other;  
Query Match 15.9%; Score 174.6; DB 20; Length 2730;  
Best Local Similarity 57.5%; Pred. NO. 1.1e-26;  
Matches 446; Conservative 0; Mismatches 244; Indels 85; Gaps 4;  
QY 2 aattcgacagagcgcatctgcttggtgctgcagcgctgaggtcgacaccagagctgc 61  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 16 aaagggccacagcgacaatgacagctgacaaaggaagaagagtagctcgagagag 75  
QY 62 ggaaggaagatcgcgagagcgcgcccgagcgagcgagcagcaggaagagagtgctgt 121  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 76 ggaaggaagatcgcgagagctgctgcggtgctcgcgagcaggaagagagagtgcttct 135  
QY 122 accagctgagcagacactctgccccttctgctgcgagcgctgacgacagctggaagagct 181



```

Db      136 atgagctggccatgagctgctctgccccacagctgtagctccatctggaagaagcct 195
Qy      182 ccatcatcgctcacaatacgaactactcgcgaatcagccgctctgcgaagc----- 235
Db      196 ccatcatcgactgagctcactctctcgcgaacacacagcctcctcctcagttgct 255
Qy      236 -----gtggaaaaaggaggagccactggaagcctgctctcctggaagcct 281
Db      256 ctgaaacagagtcgcgaagcgcgaagctgacagagatggaacactgttaccctgaagcct 315
Qy      282 tggagggttctgctgactcaccgcagagggaacatgcttaccctgctcgaagaatg 341
Db      316 tggagggttcttacttgccggtgagcccaagatgagcgcaatgtacttctcgcgaanaa 375
Qy      342 tcagcaagcaactggcctcagtcagtcgaactctgtctcctcctcctgataataacc 401
Db      376 tcagcaagctcattcgtgactataacag----- 401
Qy      402 cactctgtgtaccatctctctcctgagctcattggaacagatattgtattatca 461
Db      402 -----gtgagactaaagagacataglatcttgacttcaactca 439
Qy      462 tcctctgaccaaaggaaacttcaagaagcgcctgacccca-----ggccgaacctgac 515
Db      440 tcctctgacccatgagagatctcgtgagaaacctgagcttcaaaaagctctgtgttgg 499
Qy      516 aaagaagaagctggaagcccaacagagcgcaacttctcctggaatgaagaagcgt 575
Db      500 gaaaaaaacaaagacatgtccacagagcggaacttcttcatgagatgaagtgacagct 559
Qy      576 caccagcagaaggcgagcgcgtcaaacctcaagcgagcgctcctggaagtgctgacgctc 635
Db      560 caccacagaagcgccgaactcagctcaacctcaagtcagccacttggaagtgctgacgctc 619
Qy      636 aggaacatgagggcctcaagaagcccccctgcaagacacttccctgcccggagccctgcgc 695
Db      620 ggcccgagtgaaagctataacaacatgcctc-----ctcaaatagctgtgtggtacaaa 676
Qy      696 cgaagctccctcgcaatgcctgctgtatctatgtgaagccatcccccagctccc 750
Db      677 ggaagccctgctctcctgctcattcatctgttgaacacatccagcaccatcc 731

RESULT 11
ID      AA294052 standard; DNA; 1607 BP.
AC      AA294052;
XX      05-JUN-2000 (first entry)
DE      Human endothelial PAS domain protein-1 dominant negative mutant DNA.
XX      XX
KW      Endothelial PAS domain protein-1; EPAS1; angiogenesis;
KW      antiarteriosclerotic; antitumour; atherosclerosis; tumour;
KW      gene therapy; human; dominant negative mutant; ss.
OS      Homo sapiens.
OS      Synthetic.
XX      XX
FH      Key      Location/Qualifiers
FT      CDS      150..1610
FT      /tag= a
XX      WO200009657-A2.
XX      24-FEB-2000.
XX      13-AUG-1999; 99WO-US18539.
XX      14-AUG-1998; 98US-0096515.
XX      XX

```

```

PA      (HARD ) HARVARD COLLEGE.
XX      XX
PI      Lee M, Maemura K, Hiesh C;
XX      XX
DR      WPI; 2000-205996/18.
XX      P-PSDB; AAY79160.
XX      XX
PT      Modulation of angiogenesis in mammals, useful for treating e.g.
PT      atherosclerosis, tumors, wounds, vascular disease, hypoxic tissue
PT      damage, ischemia, balloon angioplasty, frostbite, gangrene or poor
PT      circulation
XX      XX
PS      Disclosure: Page 4-5; 57pp; English.
XX      XX
CC      The present sequence is that of DNA encoding a dominant negative
CC      mutant (see AAY79160) of human endothelial PAS domain protein-1 (EPAS1).
CC      The mutant lacks the transactivating domain (see AAY79158) of the
CC      wild-type protein (see also AAY79161). The invention is based on the
CC      discovery that EPAS1 binds to cis-acting regulatory sequences
CC      associated with genes encoding angiogenic factors such as vascular
CC      endothelial cell growth factor (VEGF) and VEGF receptors such as
CC      KDR/Flk-1 and Flt-1, thereby transactivating the promoters of such
CC      genes. A claimed method of inhibiting angiogenesis in a mammal
CC      comprises administering to the mammal a compound which inhibits
CC      binding of EPAS1 to the cis-acting transcription regulatory DNA of
CC      an angiogenic factor (see AA294051). The compound may be an EPAS1
CC      polypeptide lacking a transactivation domain or a nucleic acid
CC      encoding such a polypeptide. When such an EPAS1 mutant is bound to
CC      a cis-acting regulatory DNA, it prevents wild-type EPAS1 binding and
CC      thereby inhibits transcription of a gene encoding an angiogenic
CC      factor such as VEGF, and hence angiogenesis. The mutant is preferably
CC      administered to an atherosclerotic lesion or to a tumour site.
CC      Angiogenesis is also inhibited using a compound, such as an EPAS1
CC      dominant negative mutant, which inhibits binding of EPAS1 to the
CC      EPAS-binding element, ARNT4 (see AAY79162).
XX      XX
SQ      Sequence 1607 BP; 395 A; 480 C; 442 G; 290 T; 0 other:

Query Match      15.7%; Score 173; DB 21; Length 1607;
Best Local Similarity 57.4%; Pred. No. 2,1e-26;
Matches 445; Conservative 0; Mismatches 245; Indels 85; Gaps 4;

Qy      2 aattcgacagaggcgcatggtgtggtcgcgcgcgtgagtgtaacaccgaactgc 61
Db      133 aaagggccacagcagacaatgacagctgacaagaagaagaatgactcggaagaga 192
Qy      62 ggaaggagaagctcgcggaagcgccgcagccgcgcagcggaagcgaagagtgctgt 121
Db      193 ggaaggagaagctccggtatgctcgcggtgcccggcggaaggaagagagtgcttct 252
Qy      122 accagctgcgcacactctgccccttgcgcgcgcgtcagcgcaactggaagaagcct 181
Db      253 atgagctggccatgactcgcctctgcgccacagtgtagctccatctggacaagcgt 312
Qy      182 ccatcatcgctcacaatacgaactactcgcgaatcagccgctctgcgaagc----- 225
Db      313 ccatcatcgactggaatacagcttctcgtcgaacacacagctcctcctcagttgct 372
Qy      236 -----gtggaaaaaggaggagccactggaagcctgctctcctggaagcct 281
Db      373 ctgaaacagagtcgcgaagcgcgaagctgacagagatggaacactgttaccctgaagcct 432
Qy      282 tggagggttctgctgactcaccgcagagggaacatgcttaccctgctcgaagaatg 341
Db      433 tggagggttcttacttgccggtgagcccaagatgagcgcaatgtacttctcgcgaanaa 492
Qy      342 tcagcaagcaactggcctcagtcagtcgaactctgtctcctcctcctgataataacc 401
Db      493 tcagcaagctcattcgtgactataacag----- 518
Qy      402 cactctgtgtaccatctctcctgagctcattggaacagatattgtattatca 461

```



XX	AA294053;
AC	
XX	
DT	05-JUN-2000 (first entry)
XX	
DE	DNA encoding human endothelial PAS domain protein-1.
XX	
KW	Endothelial PAS domain protein-1; EPAS1; human; angiogenesis;
KW	antiarteriosclerotic; antitumour; atherosclerosis; tumour;
KW	vascular disease; vulnerable; carotid; vasotropic;
KW	cerebroprotective; gene therapy; ss.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Location/Qualifiers
FT	150..2762
FT	CDS
FT	/*tag= a
XX	
PN	WO200009657-A2.
XX	
PD	24-FEB-2000.
XX	
XX	13-AUG-1999; 99WO-US18539.
PF	
XX	14-AUG-1998; 98US-0096515.
PR	
XX	(HARD ) HARVARD COLLEGE.
PA	
PI	Lee M, Maemura K, Hlesh C;
XX	
DR	WPI: 2000-205996/18.
XX	P-PSDB: AAY79161.
PT	
PT	Modulation of angiogenesis in mammals, useful for treating e.g.
PT	atherosclerosis, tumors, wounds, vascular disease, hypoxic tissue
PT	damage, ischemia, balloon angioplasty, frostbite, gangrene or poor
PT	circulation -
XX	
PS	Disclosure: Page 14-15; 57pp; English.
XX	
CC	The present sequence is that of cDNA encoding human endothelial PAS
CC	domain protein-1 (EPAS1, see AAY79161). The invention is based on
CC	the discovery that EPAS1 binds to cis-acting regulatory sequences
CC	associated with genes encoding angiogenic factors such as vascular
CC	endothelial cell growth factor (VEGF) and VEGF receptors such as
CC	KDR/Flk-1 and Flt-1, thereby transactivating the promoters of such
CC	genes. A claimed method of inhibiting angiogenesis in a mammal
CC	comprises administering to the mammal a compound which inhibits
CC	binding of EPAS1 to the cis-acting transcription regulatory DNA of
CC	an angiogenic factor (see AA294051). The compound may be an EPAS1
CC	polypeptide lacking a transactivation domain (see AAY79160) or a
CC	nucleic acid encoding such a polypeptide, an antisense nucleic acid
CC	complementary to mRNA encoding EPAS1, or an EPAS1-specific antibody.
CC	The compound is preferably administered to an atherosclerotic
CC	lesion or to a tumor site. Angiogenesis is also inhibited using a
CC	compound, such as an EPAS1 dominant negative mutant, which inhibits
CC	binding of EPAS1 to the EPAS1-binding element, ARNT4 (see AAY79162).
CC	Angiogenesis can be promoted by administering EPAS1 DNA to increase
CC	expression of VEGF or VEGF receptor in endothelial cells of a
CC	patient suffering from peripheral vascular disease, cerebral
CC	vascular disease, hypoxic tissue damage (e.g. hypoxic damage to
CC	heart tissue), or coronary vascular disease as well as to treat
CC	patients who have, or have had, transient ischemic attacks,
CC	vascular graft surgery, balloon angioplasty, frostbite, gangrene,
CC	or poor circulation.
XX	
SO	Sequence 2818 BP; 669 A; 882 C; 764 G; 503 T; 0 other;

Query Match 15.7%; Score 173; DB 21; Length 2818;

Best Local Similarity 57.4%; Fred. No. 2.3e-26;

Matches 445; Conservative 0; Mismatches 245; Indels 85; Gaps 4;

[illegible]

PN WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225557.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226779.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0233297.  
PR 14-SEP-2000; 2000US-0233298.  
PR 14-SEP-2000; 2000US-0233299.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
PI Rosen CA, Barash SC, Ruben SM;



Qy 503 -----ggccgaacctgtcaagaagaagcttgaagcccccaacagagcgccactttccc 556  
Db 491 aaatgacctgtgttttggaataaagaacatgttcaacagagcggtacttctca 550  
Qy 557 tgcgaatgaagagcagcttcaaccagagagggcgacgctcaactcaagcgccacct 616  
Db 551 tgaagatgaagtgacaggtcaccaacagagcggtactgttcaacctcaagtcagccacct 610  
Qy 617 ggaaggtgtgacctgtcaagacatatgagggcttacaagccccctgtcacagacttccc 676  
Db 611 ggaaggtcttgacctgtcacagggcgagtgaaagtctacaacaactgacctc---ctacaa 667  
Qy 677 ctgcccggagccctcgtcccgagccctcccttgcgaatgacctgtgtctatctgtgaagcca 736  
Db 668 atagctgtgtgtacacaagagccccctgtctctgtcctatcatcatgtgtgaaccaa 727  
Qy 737 tccccagctcccc 750  
Db 728 tccagcaccatccc 741

Search completed: September 8, 2002, 06:49:31  
Job time: 6984 sec

OY 2 aaatcgcgcaagagggcatgtgcgttggcgctgcgaugcgttgagtgcgaacacgaagcttgc 61  
|| ||| | ||| | ||| | ||| | ||| |  
Db 133 AAAGGGCCACAGCGACATGACACTGCACAGAAGGAAGAAAGGAGTACCTTCGAGAGGA 192

164968/20

QY	62	ggaagaggaatctgcgggaacgcggcccgacggcgccacgagagacgagagatctgct	121
Db	193	ggaagaggaatctcccggaatctctgcgcgtcccgccgacgaagaaacgaagatgtct	252
QY	122	accagctggcgcacactctgcctcttcgcgcgcgcgtcaacgcgcacacttgacaagctc	181
Db	253	atgacgtgcggcccaatgacctgcctctgcgcccaactgtagactcccatctggacaagcct	312
QY	182	ccatcatgcgctctacaatgaagctacactctgcaatgacccgcctctcgcacgag----	235
Db	313	ccatcatgcgctctgaatatacgtcttcctggaaacacaaagctctctccctgacgttgc	372
QY	236	-----gtgaaaaggggagagacacactgcgcgcgtctgactaaagccc	281
Db	373	ctgaaaacgaatccggaagccgaagactgacacagacgagatggacaactttacttgaaagcct	432
QY	282	tggagggatctcgtcaatgctactcaacgcgcgagagagacatgcttacctctggaata	341
Db	433	tggagggatctcgtcaatgctactcaacgcgcgagagacatgcttacctctggaataac	492
QY	342	tcaagaagcaactctgcgcctcaagtaagtgagaaactctgctccctccctgtacataacc	401
Db	493	tcacgaacttcatatggacttaccacg-----	518
QY	402	cactccctggtacaaattctctctgcgaagctcaatgacacagtaactcttgatttacc	461
Db	519	-----gtggagactaacagagacacatgatatctttgacttccactca	556
QY	462	tccctctgacccaagaagaaacttcaagaagcgtccgaacccca-----gscgcgaactgtc	515
Db	557	ttccctgcaccatatgagagatctctgcgaacactgaagctcaaaaatggcctctgctttgg	616
QY	516	aaagaagaagcttgaagccccaagaagagcgcaacttctccctcgaatgaagaagcagct	575
Db	617	gaaaaaaaagcamaagacatgtgtccacagacggagactcttcatatgagatgaagacgggt	676
QY	576	caccaagaagaagggcgacgcgtccaaactcaagaagcgagcaacttgaaagtgtgtacatgtc	635
Db	677	caccaagaagaagccgtatgtctcaacctcaaaagtcacgcccaactggaaaggtcttggacatgcac	736
QY	636	aggaacatatggggcctcaacaagccccctgcacagacttccctctgcgcggaagccctcgctc	695
Db	737	gggcgcagctgaaatcttacaacacactgcctc---ctcacatatagttgctgtggatracaa	793
QY	696	cgaaagctcccccgtcaatgctcgcgtgtatctgtgaaagccatcccccagctccc	750
Db	794	ggagccctctgctctgctccctcatcatcatatgtgtgaacacatccagacaccatcc	848

RESULT 2  
 US-08-785-241-2  
 : Sequence 2, Application US/08785241  
 : Patent No. 5695963  
 : GENERAL INFORMATION:  
 : APPLICANT: McKnight, Steven L.  
 : APPLICANT: Russell, David W.  
 : APPLICANT: Tian, Hui  
 : TITLE OF INVENTION: Endothelial PAS Domain Protein  
 : NUMBER OF SEQUENCES: 7  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
 : STREET: 268 BUSH STREET, SUITE 3200  
 : CITY: SAN FRANCISCO  
 : STATE: CALIFORNIA  
 : COUNTRY: USA  
 : ZIP: 94104  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC COMPATIBLE  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:

```

? APPLICATION NUMBER: US/08/785,241
? FILING DATE: 17-JAN-1997
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: OSMAN, RICHARD A.
? REGISTRATION NUMBER: 36,627
? REFERENCE/DOCKET NUMBER: UTSD:122
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 343-4341
? TELEFAX: (415) 343-4342
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3031 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
?
US-08-785-241-2

```

Query Match	14.28;	Score	156.6;	DB 1;	Length	3031;
Best Local Similarity	57.18;	Pred. No.	6e-28;			
Matches 441;	Conservative	0;	Mismatches	244;	Indels	88;
					Gaps	5

[illegible]



Db 749 ACCCCCTGCTGCTGCTTATCATCATGTGTGAGCCATTCAGACCCATCC 801

RESULT 3  
US-08-480-473B-1  
; Sequence 1, Application US/08480473B  
; Patent No. 5882914

GENERAL INFORMATION:

APPLICANT: Semenza, Gregg L.

TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/480,473B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 07265/053001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619/678-5070

TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3736 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-480-473B-1

Query Match 12.2%; Score 134.6; DB 2; Length 3736;  
Best Local Similarity 54.7%; Pred. No. 1e-22;  
Matches 387; Conservative 0; Mismatches 244; Indels 76; Gaps 3;

QY 53 ccgaagctgaggaaggaagtcgcggagcgccgcgagccgcgagccaggaagcag 112  
Db 72 CTGAACGCTGAAAAGAAAGCTCGAGATCGAGCAGCCGATCTCGGCAATAAAGATCTG 131  
QY 113 aggtctgtacacagctgcgcacacctctgccttgcgcgcgcgcgcgcgcgccttg 172  
Db 132 AAGTTTATATAGCTTGTCTCATCTCCACTTCCACATATATGTGAGTTGCCATCTTG 191  
QY 173 acaagcctcactcactcgcctcacaatcactctgcacatgcacccgcctctgcag 232  
Db 192 ATAAAGCCCTCTGTGATGAGCTTACCATCAGATATTTGGTGAGAAACTCTTGAG 251  
QY 233 caggtg-----gaaaaagggagagccactgtgaagcctctactcgaag 278  
Db 252 CTGGTGATTGTGATATTTGAATGATGACATGAAAGCAGATGAAATGCTTTATTTTAAAG 311  
QY 279 ccctgagaggttctgcctgactcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 338  
Db 312 CTTGGATGTTTGTATGATGTTCTTCTACAGATGATGATGATGATGATGATGATGATGAT 371  
QY 339 atgtcaagcagcctgcagctcagctcagctgtgacctgtctcctccctgatacataa 398  
Db 372 AATGTAACAAATACATGAGATTAACCTCACT----- 401

QY 399 cccactcgtgtaccattctctcgtgagctcacttgagacagatcttattat 458  
Db 402 -----TTGACTACTGGACACAGTGTGTTGATTTTAC 435  
QY 459 ccattcctgtgacccaaggaactcagaagcgccttaccgcccccgcgaacctgtcaaa 518  
Db 436 TATCCATGTGACATGAGGAATGAGAAATAGCTTTACACACAGAAATAGCCTTGTA 495  
QY 519 gaagaagctggaagccccaagagcgcacttctccctgcgaatgaagacgcgtcac 578  
Db 496 AAAGGTTAAAGAACAAACACACAGGAGCTTTTCTCAGATATAGTTACCTTAC 555  
QY 579 cagcaagagggcgacgctcactcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 638  
Db 556 TACCCAGAGAACTATGATTAAGTCTGCAATGAGGATATTCACATGCGACAG 615  
QY 639 aatatgagggcctacaagccctgcagagacttccctgcgcgcgcgcgcgcgcgcgc 698  
Db 616 CCACATTCACGATATGATACCA-----ACACTAACCAACCTCAGTGGGTATAGAA 669  
QY 699 gcttccctgcgaatgctgtgtatctgtgaagcattcccccaag 745  
Db 670 ACCACCTATGACCTGCTGTGCTGATTTGTGAACCATTCCTCACC 716

RESULT 4  
US-08-915-213-1  
; Sequence 1, Application US/08915213  
; Patent No. 6020462

GENERAL INFORMATION:

APPLICANT: Semenza, Gregg L.

TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/915,213

FILING DATE: 20-AUG-1997

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/480,473

FILING DATE: 06-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 07265/053001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619/678-5070

TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3736 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-915-213-1

Query Match 12.2%; Score 134.6; DB 3; Length 3736;  
Best Local Similarity 54.7%; Pred. No. 1e-22;  
Matches 387; Conservative 0; Mismatches 244; Indels 76; Gaps 3;

OY	53	ccgagctcggaagaagaatctcgcggaacgcgcccgcaagcccggaacgaagaacg	112
OY	53	ccgagctcggaagaagaatctcgcggaacgcgcccgcaagcccggaacgaagaacg	112
Db	72	CGGAACGTCGAAAAGAAAGTCTCGAATGTCAGCCAAATCTCGCGAAGTAAAGATCTG	131
OY	113	aggctcgtacacagctgagcgacacctctgcctcttgccgcgcgcgtacgcgcacctg	172
Db	133	AGGTTTTATAGCTTGCTGCTATCAGTTGCCACTTCCACATAAATGATGATTCGATCTTG	191
OY	173	acaagagctccatcatgctgctccaaatcagctccacgcgcgcgtacgcgcacctg	232
Db	192	ATTAAGCCCTCTGTGTAGAGGCTTACCATCACCTATTGTCGGTGAAGAAACTTCGTGATG	251
OY	233	caagtg-----gaaaaggggagagccacgtggaacgcgcctgtctactacga	278
Db	252	CGGGGATTTGGATATTGACATGACATGAAGAAGCAGAGTAATGCTTTTATTTGAAAC	311
OY	279	ccctgagaggtcttgctcatgtaactaccgcgcgagagagatgcttaccctgcgaa	338
Db	312	CCCTTGGATGGTTTTGTATTGGTTCTCACAGATGATGGTGACATGATTTCATTTCGATA	371
OY	339	atgtcagaagaacgcttgagcgctcgaatgtagtgaactctgttccctccctgatactaa	398
Db	372	ATGTGAACAAATACATGGGATTAATCTCAGT-----401	401
OY	399	ccccactcctgylaccaattctctctgagctcatatggaacaaagtaactcttgaattat	458
Db	402	-----TTGAACTAATCGACAGACAGTGTTGTTGATTTTAC	435
OY	439	ccatccctgtaccgaaggggaacttaagaagccctgtaccccccgacgcgcgaactgtcaa	518
Db	436	TCATCATGTAATACCATGTAGGAAATGAGAAATCTTACACAGAAATGGCCTTGTA	495
OY	519	gaagaagcttggaagccccaacagagcgcccaactttccctcgcgaatgaaagagacgctcac	578
Db	496	AAAGGGTAAAGAACAAAACACACAGCAAGCTTTTCTCGAATGAAAGTACCTAAC	555
OY	579	cagcagagggcgcaagcttaaacctcaaaagcgcccaactggaagtgctgcaactgtcag	638
Db	556	TAGCCGAGGAAGACTAATGAACATAAAGTCTGCACATGAGAGTATTGCCATGCGACACG	615
OY	639	acatctgagggcctcaacagccctctgcaaggaactccctccgcgcgcgcctctgcctga	698
Db	616	CCACATTACAGTATTATGTATACCA-----ACAGTAACCAACTTCAGTGTGGTATTAAGA	669
OY	699	gcctccctctgcaatgctgtgtcttactctgaaagccatcccccagc	745
Db	670	ACCACCTATGACCTGGCTGGTGTGATTTGAAACCCATTCTCTACG	716
RESULT 5			
US-09-148-547-1			
; Sequence 1, Application US/09148547			
; Patent No. 6124131			
; GENERAL INFORMATION:			
; APPLICANT: Semenza, Gregg L.			
; TITLE OF INVENTION: Hypoxia Inducible Factor-1 and Methods of Use			
; FILE REFERENCE: 07265/151001			
; CURRENT APPLICATION NUMBER: US/09/148,547			
; CURRENT FILING DATE: 1998-08-25			
; NUMBER OF SEQ ID NOS: 2			
; SOFTWARE: patentlin Ver. 2.0			
; SEQ ID NO 1			
; LENGTH: 3736			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (23)..(2509)			
; US-09-148-547-1			

Query Match	12.28;	Score 134.6;	DB 3;	Length 37336;
Best Local Similarity	54.7%;	Pred. No. 1e-22;		

Matches	387;	Conservative	0;	Mismatches	244;	Indels	76;	Gaps	3
QY	53	ccgagctcggaaggaagatcgcgagacgagcccgacgacgagcaaggaacgg	112						
Db	72	ctgaacgtcgaaaaagaagctcgtgatgcagccagatctcgcgaaatgaagatctg	131						
QY	113	aggtgtgttaccagctcgagcacaactctgaccttggcgcgggcgttcaagcgcaacttgg	172						
Db	132	aagttttttatgtagcttgcctcatcagttgcaccttccacaataatgtgtgttcgcaccttgg	191						
QY	173	acaagaccctcatcgtcgcctccacaatcagcttacctgtgcgaatgcacgacctgtcgacg	232						
Db	192	ataagggccctctgtatgatgaagcttaccatcagtatttggctgttgagaaacttctggatg	251						
QY	233	casgtgtg-----gaaaagggggagagccacattgagcgcgtgtctaacttgaag	278						
Db	252	ctgtgtatttgtatattgaaatgatgcatagtaaagcacaagatgatgttatttattgaag	311						
QY	279	cctctgaaaggtttctgtcattgtaactcaccgcggaaggaaactgtgcttacctgtcgaa	338						
Db	312	cccttgatgtgttttggcttgaagcttccacgaagatgtgtacatgtatttcaattcttgata	371						
QY	339	atgtgcagaagacacactcgggccctcagctcgtggaacctgtctctcctccgtatacata	398						
Db	372	atgtgaacaataatcatatggttaactcagtt-----	401						
QY	399	cccacactcctgtgtaccaatcttctctctgtgagctcatcttgacacagtatcttggatttat	458						
Db	402	-----ttgaactaactggaacagtggtgtgattttac	435						
QY	459	ccatccctgtgaccaaaggaagaaacttcaagaagcccttgaccccgacgacgttcaaa	518						
Db	436	tatctcatgtgcacatgagaaatgagaaatgtcttacaacaaagaatgacctgttgaa	495						
QY	519	ggaagaagctcggaagccccaacagagcgccaactttccctggcggaaggaagagacgttacc	578						
Db	496	aaaagggtaaagaaacaaacacacagcgaagccttttctccaaatgaaatgtatccctaac	555						
QY	579	caagcagaagcgacaagctcaacctcaaaagcggccacacttggaaagtgctgcacatgtccag	638						
Db	556	tagccgaggaagaacatctgaacataaagctctgcaacatgtgaagtgatttgcacttgacaag	615						
QY	639	acatatgagggcctacaagccccctgcacagaacttcccctgcgggagacctcgtctcga	698						
Db	616	ccaacattccagtatatgtatacca-----acagtaacaacaacctcagtgtygtgataagaa	669						
QY	699	ggctccctcgcaatgcctgtgtgtcttatctgttataagacatccccacg	745						
Db	670	accacctatgacctgtcgtgtgtatctgtgaatgtgaacccattctctacc	716						

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,217
;

```

FILED DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,473  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/053001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3736 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-09-235-217-1

Query Match 12.2%; Score 134.6; DB 4; Length 3736;  
Best Local Similarity 54.7%; Pred. No. 1e-22;  
Matches 387; Conservative 0; Mismatches 244; Indels 76; Gaps 3;

QY 53 ccgagctcgcggaaggaagatcgcggaagcccgccgagcccgcgagcccgaggaagcgg 112  
DB 72 CTGAACGTCGAAAGAAAGTCTCGAGATGCGAGCCAGATCTCGGGGAGTAAGATCTG 131  
QY 113 aggtcctctaccagctgctgcgcacactctgccccttgccgcgcgctcagcgccactctg 172  
DB 132 AAGTTTTTATAGCTTGCTGCTCATGCTGCCACTTCCACATATATGATGACTTGCCATCTTG 191  
QY 173 acaagcgctccatcatctgcgtccacatcagctacactgcatctgcaatgcgcgctctgcgag 232  
DB 192 ATAAGCCCTCTGTGATGAGGCTTACCATCAGCTATTGGGTGTGAGGAACCTTCGAGATG 251  
QY 233 caggtg-----gaaaaaggggagagccactgtagcgctgctactcgtgaag 278  
DB 232 CTGGGATTTGGATATTGAAGATGACATGAAGACAGATGATTTGCTTTTGTGAAG 311  
QY 279 cccctgaggggttcgtcatctgctacacgcgcgagggagagacatgcttaccgtcgagaa 338  
DB 312 CCTTGATGAGTGTGTTTGTATGTTCTCAGATGATGATGATGATGATGATGATGATGATGATG 371  
QY 339 atgtcaagaagcactggtgctcagctcagtggaactctgttccctccctcctgatacataa 398  
DB 372 ATGTGAACAATATACATGGGATTAAGTCACTG----- 401  
QY 399 ccccaactcctgtaaccaatttctctctgagctcatgtggaacagatatttgattttat 458  
DB 402 -----TTGAACTAACTGGACACAGTGTGTTGATTTTAC 435  
QY 459 ccatccctgtgacaaaggaagaaactcacaagcccttgaaagcccccagagccgaactgtcaaa 518  
DB 436 TTAATCATATGACCAATGAGGAATATGAGCAATATGCTTACACACAGAAATGGCTTGTGAA 495  
QY 519 gaagaagctggaagcccaacagagcgccaactttccctcgaaatgaagagcagctcac 578  
DB 496 AAGGGTAAAGAAACAAACACACAGGAGCTTTTTCAGAAATGAAGTGAAGTGAAGTGAAGTGAAG 555  
QY 579 cagcagagggcgagctcaacactcaagcgagcagcttgaagggcgctgcatctgctcag 638  
DB 556 TTAGCGAGGAACATATGAAACATAAAGCTGCACATGGAAGGTATGTCACCTGCACAGG 615  
QY 639 acaatgaagcgctcaaaagcccccgcacagagactccctcgcgagagccctgcgcag 698  
DB 616 CCACATTCACGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 669  
QY 699 gctccctcgtcaatgctggtgcttactctgtgaagccatccccagc 745  
DB 670 ACCACTTATGACCTGCTGTGCTGATTTGTGAACCATTCCTCAAC 716

RESULT 7  
PCT-US96-10251-1  
Sequence 1, Application PC/TUS9610251

GENERAL INFORMATION:  
APPLICANT: The Johns Hopkins University School of Medicine  
TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10251  
FILING DATE: 06-JUN-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/053001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3736 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
PCT-US96-10251-1

Query Match 12.2%; Score 134.6; DB 5; Length 3736;  
Best Local Similarity 54.7%; Pred. No. 1e-22;  
Matches 387; Conservative 0; Mismatches 244; Indels 76; Gaps 3;

QY 53 ccgagctcgcggaaggaagatcgcggaagcccgccgagcccgcgagcccgaggaagcgg 112  
DB 72 CTGAACGTCGAAAGAAAGTCTCGAGATGCGAGCCAGATCTCGGGGAGTAAGATCTG 131  
QY 113 aggtcctctaccagctgctgcgcacactctgccccttgccgcgcgctcagcgccactctg 172  
DB 132 AAGTTTTTATAGCTTGCTGCTCATGCTGCCACTTCCACATATATGATGACTTGCCATCTTG 191  
QY 173 acaagcgctccatcatctgcgtccacatcagctacactgcatctgcaatgcgcgctctgcgag 232  
DB 192 ATAAGCCCTCTGTGATGAGGCTTACCATCAGCTATTGGGTGTGAGGAACCTTCGAGATG 251  
QY 233 caggtg-----gaaaaaggggagagccactgtagcgctgctactcgtgaag 278  
DB 252 CTGGGATTTGGATATTGAAGATGACATGAAGACAGATGATTTGCTTTTGTGAAG 311  
QY 279 cccctgaggggttcgtcatctgctacacgcgcgagggagagacatgcttaccgtcgagaa 338  
DB 312 CCTTGATGAGTGTGTTTGTATGTTCTCAGATGATGATGATGATGATGATGATGATGATGATG 371  
QY 339 atgtcaagaagcactggtgctcagctcagtggaactctgttccctccctcctgatacataa 398  
DB 372 ATGTGAACAATATACATGGGATTAAGTCACTG----- 401  
QY 399 ccccaactcctgtaaccaatttctctctgagctcatgtggaacagatatttgattttat 458  
DB 402 -----TTGAACATACTGGACACAGTGTGTTGATTTTAC 435

Accession	Sequence	Position
Oy	ccatcccgctgtacccaagaagagacttaagaagccctgtgacccccagcgccgaacctglttaa	518
Oy	459	
Db	TCATCATGTAATGACCTGTGAGGAATGAGAAATGCTTTCACACAGGAATGAGCTTGAA	495
Db	436	
Oy	gaagaagcttgaagccccaagaagcgccgaactttctccctcgaaatgaagaacagctcac	578
Oy	519	
Db	AAAGGGTAAAGAAACAAACACACGCAAGCTTTTTCTCGAGATGAAGTGAACCTTAC	555
Db	496	
Oy	cagcgaagcgcgacgcgtcaaacctcaagaagcgccgaacttgaagtgctgacgtctag	638
Oy	579	
Db	TAGCCGAGGAAGAACTATGAACATTAAGTCTGCACATAGGAAGTATTTGCACTGCACAGG	615
Db	556	
Oy	acatatgaagggcctctcaaacgccccctgcagagacttccctcgccggagacccctgcgtcca	698
Oy	639	
Db	CCACATTCACCTATATGATATACCA-----ACAGTAACCAACCTCAGTGTGGGTATTAAGAA	669
Db	616	
Oy	gcttcacctgcaatgctcgtggtgtctatctgtgaagccatccccagc	745
Oy	659	
Db	ACCACTATATGACCTGCTTGCGTGATTTTGGAACCCATTCCTTAC	716
Db	670	

```

      RESULT 8
US-08-785-310A-1
; Sequence 1, Application US/08785310A
; Patent No. 5840532
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L.
APPLICANT: Russell, David W.
TITLE OF INVENTION: Neuronal PAS Domain Protein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,310A
FILING DATE: 21-JAN-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UTSJ:1226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2078 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IS-08-785-310A-1

```

Query Match	8.1%;	Score 89.6;	DB 2;	Length 2078;
Best Local Similarity	67.98;	Pred. No. 3.3e-12;		
Matches 125;	Conservative 0;	Mismatches 59;	Indels 0;	Gaps 0;

[illegible]

Db 370 GAGTCTTCGAGCTGGGGCAACTTCTCCCGCTGGCCGGCGCATCTCCATTCAGCTGAC 429

Oy 175 aaggtccatcaltatgcgcctcacaataagctactgtgcabgcacgcgcctctgcagca 234

Db 430 AAGCTTCCATCTGTCGGCTCAGTGACCTACCTCCGCTGCGCCCGTTCGCGCGCTG 489

Oy 235 ggtg 238

Db 490 GGGG 493

```

US-08-785-310A-2
US-08-785-310A-9
Sequence 2, Application US/08785310A
Patent No. 5840532
GENERAL INFORMATION:
APPLICANT: McNight, Steven L.
TITLE OF INVENTION: Neuronal PAS Domain Protein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,310A
FILING DATE: 21-Jan-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UTSD:1226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2082 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: Linear
MOLECULE TYPE: CDNA
US-08-785-310A-2

```

Query Match	6.58;	Score 72;	DB 2;	Length 2082;
Best Local Similarity	62.08;	Pred. No. 4.7e-08;		
Matches 114;	Conservative 0;	Mismatches 70;	Indels 0;	Gaps 0;

QY	55	gagctctggaagggaaagctgcgsggaagcgagccgagccgagcgagcgagcgag	11.4
	190	GGGAGGGCAAGAAAGTCCCGAAGCTGCGCGGCGAGCCGGGAAGAAATCTG	24.9
QY	115	gtctgtaccagctcgcgacacctctgcctcttgcgcgcgctcagcgcacacttgac	17.4
Db	250	GAGTCTTTCGAGCTAAGCCAAAGCTCTCTCCCTCTGCCCCGTGCATCTTCAAGCAAGCTGGAC	30.9
QY	175	aaggcctccatcatcgcgcctccacatcagcttacccttcgcatgacgcgcctctcgagca	23.4
Db	310	AAGCGTCATCGTGGCTTTAAGGTCACTCACTCCGCTGGCTGGCTTTTGGCGGCGCTG	36.9
QY	235	ggtg	23.8
Db	370	GGGG	37.3

```

TITLE OF INVENTION: Process for Its Production
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/ms-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738,349
FILING DATE: 25-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/364,439
FILING DATE:
APPLICATION NUMBER: US 08/112,061
FILING DATE: 26-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Barker, M. P.
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 02481.1323-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3581 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Mus musculus
STRAIN: osteoblastic cell line MC3T3E1
FEATURE:
NAME/KEY: CDS
LOCATION: 284..2671
US-08-738-349-1
Query Match. 5.7%; Score 62.8; DB 2; Length 3581;
Best Local Similarity 81.1%; Pred.No.8.6e+06;
Matches 73; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
OY 1011 aaagacctcctccgttgtaggcaccagagaaaaaa 1070
| | | | | | | | | | | | | | | | | | | | | |
Db 3486 AAAAAAATCTTTTAAAAAAAAAAAAAAAAAAAAAA 3545
OY 1071 aaaaaaaaaaaaaaaaaaacatgaggccgc 1100
| | | | | | | | | | | | | | | | | | | | | |
Db 3546 AAAAAAAAAAAAAAAAAAGCGCCGC 3575
RESULT 12
US-08-722-126A-4
Sequence 4, Application US/08722126A
Patent No. 6034227
GENERAL INFORMATION:
APPLICANT: PECHT, Israel
APPLICANT: GUTHMANN, Marcelo D.
APPLICANT: TAL, Michael
TITLE OF INVENTION: A DNA MOLECULE ENCODING A MAST CELL
TITLE OF INVENTION: FUNCTION-ASSOCIATED ANTIGEN (MAFA)
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
```

```

: STREET: 419 Seventh Street N.W., Ste. 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 20004
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/722,126A
: FILING DATE: 08-OCT-1996
: CLASSIFICATION: 536
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04258
: FILING DATE: 06-APR-1995
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: IL 109257
: FILING DATE: 08-APR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: PECHT-1A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 737-3528
: TELEFAX: (202) 628-5197
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1461 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 54..617
: US-08-722-126A-4
:
: Query Match 5.7%; Score 62.2; DB 3; Length 1461;
: Best Local Similarity 71.3%; Pred. No. 8.4e-06;
: Matches 82; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
:
: Qy 977 ccagaaaccacaatgtctcaaaaccacataaagaccctctgttgtagccag 1036
: Db 1343 CCTTAAATTTAAAGAGTCCCATTTAAATTAATATCCCTTTCGCTAAAAA 1402
:
: Qy 1037 agaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 1091
: Db 1403 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1457
:
: RESULT 13
: PCT-US95-04258-4
: Sequence 4, Application PC/TUS9504258
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: A DNA MOLECULE ENCODING A MAST CELL
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04258
: FILING DATE: 06-APR-1995
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: IL 109257
: FILING DATE: 08-APR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: PECHT-1 PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1461 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 54..617
: PCT-US95-04258-4
:
: Query Match 5.7%; Score 62.2; DB 5; Length 1461;
: Best Local Similarity 71.3%; Pred. No. 8.4e-06;
: Matches 82; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
:
: Qy 977 ccagaaaccacaatgtctcaaaaccacataaagaccctctgttgtagccag 1036
: Db 1343 CCTTAAATTTAAAGAGTCCCATTTAAATTAATATCCCTTTCGCTAAAAA 1402
:
: Qy 1037 agaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 1091
: Db 1403 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1457
:
: RESULT 14
: US-09-227-357-71
: Sequence 71, Application US/09227357
: Patent No. 6342581
: GENERAL INFORMATION:
: APPLICANT: Fischer et al.
: TITLE OF INVENTION: 123 Human Secreted Proteins
: FILE REFERENCE: P2010P1
: CURRENT APPLICATION NUMBER: US/09/227,357
: CURRENT FILING DATE: 1999-01-08
: EARLIER APPLICATION NUMBER: PCT/US98/13684
: EARLIER FILING DATE: 1998-07-07
: EARLIER APPLICATION NUMBER: 60/051,926
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/052,793
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,925
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,929
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/052,803
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/052,732
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,931
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,932
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,916
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,930
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,918

```



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 04:03:27 : Search time 1715.42 Seconds  
(without alignments)  
8654.820 Million cell updates/sec

Title: US-09-896-791b-2

Perfect score: 1100

Sequence: 1 gaattcgagcagcgccat.....aaaaaaaaacatcgcgccgc 1100

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estov:\*  
5: em\_estov:\*  
6: em\_estov:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	507.6	46.1	687	10	BF722881 mab24h10.
2	461	41.9	461	9	AL322407 m119b08.y
3	449.6	40.9	905	9	AL528423 AL528423
4	444.8	40.4	856	9	AL519496 AL519496
5	427	38.8	919	9	AL541807 AL541807
6	408.2	37.1	767	10	BG696633 BG696633
7	392.8	35.7	756	9	AU118510 AU118510
8	386.6	35.1	460	9	AA028416 AU028416
9	336	30.5	633	10	BI772109 BI772109
10	299.6	27.2	459	9	AA041707 m102902.r
11	298.2	27.1	418	10	BM118906 BM118906
12	298.2	27.1	425	10	BM118909 BM118909
13	298.2	27.1	610	10	BM119659 BM119659
14	276.8	25.2	396	10	BF653212 BF653212
15	253.2	23.0	873	10	AU116888 AU116888
16	250.2	22.7	620	10	BF365142 BF365142
17	244	22.2	865	9	AL535689 AL535689

c	18	242.8	22.1	417	9	AA988550	UI-R-CO-1
c	19	242.8	22.1	486	9	AT137683	AT137683 UI-R-CO-h
c	20	239.8	21.7	421	9	AM520329	UI-R-BUOP
c	21	239	21.8	321	9	AT171316	AT171316 UI-R-AF1-
c	22	233.2	21.2	279	9	AV322372	AV322372 AV322372
c	23	226.2	20.6	384	9	AT578946	UI-R-AAO-
c	24	211.6	19.2	733	9	AU118844	AU118844
c	25	196.8	17.9	353	9	AT510204	AT510204 m102902.y
c	26	191	17.4	430	10	BF654293	BF654293 278403 MA
c	27	188.6	17.1	228	10	BM483179	BM483179 536412 MA
c	28	174.6	15.9	858	9	AL544018	AL544018
c	29	153.4	13.9	469	10	BF548129	BF548129 UI-R-AO-b
c	30	150.2	13.7	632	9	AM133666	AM133666 f110c03.y
c	31	141.2	12.8	662	10	BM182143	BM182143 f54b05.y
c	32	139.8	12.7	1657	11	AK017853	AK017853 Mus muscu
c	33	136.6	12.4	549	10	BE721988	BE721988 189855 MA
c	34	136	12.4	659	10	BM182886	BM182886 f62h07.y
c	35	136	12.4	905	9	AL549361	AL549361
c	36	133.6	12.1	878	9	AU133294	AU133294
c	37	132.4	12.0	769	10	BG772697	BG772697 602720844
c	38	128.4	11.7	430	9	AM211226	AM211226 u129b12.y
c	39	124.8	11.3	678	9	AL642308	AL642308
c	40	123.6	11.2	533	9	AT1877949	AT1877949 fcs5e08.y
c	41	123.6	11.2	642	10	BM157049	BM157049 f43q02.y
c	42	123.6	11.2	758	10	BI890667	BI890667 2f637-2-0
c	43	122.6	11.1	526	10	BG884123	BG884123 f072b10.y
c	44	120.2	10.9	959	9	AL530209	AL530209
c	45	119.6	10.9	833	9	AL532246	AL532246

#### ALIGNMENTS

RESULT 1  
LOCUS BF722881 687 bp mRNA linear EST 03-JAN-2001  
DEFINITION mab24h10.y1 Soares.NMEBA branchial arch Mus musculus cDNA clone  
IMAGE:3971562 5' similar to TR:Q9QX54 Q9QX54 HYPKXIA-INDUCIBLE  
FACTOR 3 ALPHA ;, mRNA sequence.  
ACCESSION BF722881.1 GI:12023883  
VERSION BF722881  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 687)  
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Other\_ESTs: mab24h10.x1  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonalço  
Ph.D.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
image.lnl.gov/image/html/lresources.shtml  
MGI:1471594  
Seq primer: -40RP from Glibco  
High quality sequence stop: 469.  
Location/Qualifiers  
1 . 687  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3971562"  
/clone\_id="Soares.NMEBA\_branchial\_arch"  
/tissue\_type="branchial arches"  
/dev\_stage="embryo, 10.5 dpc"

```

/label="DH10B (phage resistant)"
/vector="pUT73D-Pac (Pharmacia)" with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
was primed with a Not I - 0190(dt) primer [5',
TGTATCCCAATCTGAAGTGGAGCGCGCCATGCTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pUT73 vector. Library
constructed and normalized by Bento Soares and M.Patima
Bonaldo."

```

Query Match	46.1%	Score 507.6	DB 10	length 687
Best Local Similarity	88.3%	Pred. No. 6.8e-54		
Matches 603	Conservative	0	Mismatches 10	Indels 70
				Gaps 2

**DY**   82 gcgcgccgacagccgcgcagcaccagagaagtagtctgtaccacagtggcgcacactctg 141  
**Db**     1 GCGGCCGACAGCCGGGCACGACGAGACGGAGTGTCTAACACAGTTGGCGCACACTCTG 60

Qy 142 ccctttgcgcgcgcgcgtcaagcgcgaacctgtgaaagcctccatcatgcccacaatc 201b  
|||||  
Db 61 CCCCTTGGCGGGCGGCGTCAGCGCGCACCTGGACAAGGCGCTCCATCATGCGGCTCACATC 120b

```
Oy      202 agctacactgcgatgcaaccgccctctgcgcag-----cagtgtgaataagg    247
          |||||
Db       121 AGCTACACTGCGCATGCACCGCCTCTGGCGAGCAGGGGAGTGAACCACAGGTGTAATAAAGG    180
```

**Oy**    248 ggagagccactgtagcgcctgtacctaaaggccctggagagtctcgcatggtactaac 307  
|||||  
|||||  
**Db**    181 GGAGGCCACTGGACGCTGCtAACCtGAAGCCCTGGAGGCTTTCATGTACTACc 240  
|||||  
|||||

QY 308 gcccaaggagacatgcttaccgtgcgaaatgtcagcaagcacctgagccctcagcag 367

Db 241 gccagggagacatgcttacctgtcgaaatgtcagcagcacctggccctcagcag 3000

Oy	368	tggaactctgtctcctccctgatacaaaccaactcctgtgaacaaattctcttg	427	ctgg	
Db	301	-----	-----	CTGG	304

Qy 428 agccattgacacaglatcttgatttatccatccctgtgacaaagaactcaag 487  
|||||  
Db 305 AGGCATTGACACACAGTATCTTTGATTTTATCCATCCCTGTGACCAAGAGGAACTTCAAG 364

Qy 488 acgccttgacccccagcgccgaaactgtcaagaagaagctgtgaagccccaacagagcgc 547  
|||  
Db 365 acgcccTGACCCCCAGGCCGAACCTGTCAAGAAGAGAGCTGGAGGCCCAACAGAGCGCC 424

[illegible]

QY 608 cggccacctggaagtgtgtcaactgtctcagacatatgagcgctacaagccccctgcaac 667  
|||||  
Db 485 CGGCCACCTTGGAGGCTGCTGCATCTGCTTCACGAGCATATGAAGGCCCTACAAGCCCTTCAC 544  
|||||

QY 668 agattccctgcgcggagccctcgctccgagctccctcgcaatgctgtgtatct 727  
db 5/5 AGACTTCCCTGCCGCGAGCCCTGCTCCGAGGCTCCCTCGCAAGCCTGGGTCCTTCT 604  
|||||

Qy 728 gtgaagccatccccagctccc 750  
|||||  
Db 605 GTGAAGCCATCCCNFACCCAGCC 627

RESULT 2  
AT322407

LOCUS	461 bp	mRNA	linear	EST 23-DEC
DEFINITION	m19b08.y1 Scores mouse p3NM19.5 Mus musculus cDNA clone			
IMAGE:463959	5' similar to TR:008787 008787 ENDOTHELIAL PAS DC			
PROTEIN 1	mRNA sequence			

ACCESSION	AI322407	
VERSION	AI322407.1	GI:4056836
KEYWORDS	EST.	
SOURCE	house mouse,	

ORGANISM

REFERENCE

AUTHORS

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE	JOURNAL	COMMENT
The Washu-HHMI Mouse EST Project	Unpublished (1996)	Contact: Marra M/Mouse EST Project Washu-HHMI Mouse EST Project

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:277775

This read is a RESEQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the  
correct orientation)  
Seq primer: -40RP from Glibco

FEATURES	High quality sequence stop: 443
source	Location/Qualifiers
	1. .461
	/organism="Mus musculus"

```

/db_xref="taxon:10090"
/clone="IMAGE:463959"
/clone_lib="Scars mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"

```

```

/lab_host="Dh10B (ampicillin resistant)"
/notes=Vector: pTR3D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'

```

TGTTACCAATCTCATAGTGGAGCGGCCGCATTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pI773 vector

(Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bent Soares and M. Fatima Bonaído. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

BASE COUNT	93 a	145 c	136 g	87 t
ORIGIN				

Query Match	41.9%	Score 461	DB 9	length 461
Best Local Similarity	100.0%	Pred. No. 4.1e-48		
Matches 461	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY 16 gccatgagcgtttggggctgcaagcgcgtgagtcgaaacaccagcctgcgaaaggagaatctcg 75  
|||||  
Db 1 gccatgagcgtttggggctgcaagcgcgtgagtcgaaacaccagcctgcgaaaggagaatctcg 60

QY 76 cggagacgcgcgcgcagccgcgcgcagccagagacgcgagtgctgtaccagctlbgcgcac 135  
|||||  
D6 61 cggagacgcgcgcgcagccgcgcgcagccagagacgcgagtgctgtaccagctlbgcgcac 120

QY 136 acctctgcccctttgcgcgcgcgtcaagcgcaccttgacaagcgcctcatcatgagctc 195  
|||||  
121 accttgcaccttttgccgcgcgcgcgtcaagcgcgaccttgacaagcgcctcatcatgagctc 180  
Db

**Dy**    acatcaagctactcgtgcgcatggaccgcgctctgtcgcagcaagtgtgaanaaggcgtagacc    255  
|||||  
**Dd**    196 acatcaagctactcgtgcgcatggaccgcgctctgtcgcagcaagtgtgaanaaggcgtagacc    255  
|||||  
      181 ACATCAAGCTACTCGTGCGCATCCACGGGCTCTGCCCACAGTAGTGAAAAAAGGGSAGAGCC    240

Oy	256	actggagccctgcgtacaccctcgaagaagcccttggaaggttttcgtaatgtagtaccacgcgcgaagg	315
Dd	241	ACTGAGCCCTGCCTACCTCGAAGAAGCCCTGAAGAGTTTCGTCAATGGTACACCGCCGAGAGG	300
Oy	316	agacaatgacctacgtctcgaaatatgtcatgaagaacgaaccttgagccctcaagtgcacctc	375
Dd	301	AGACAATGGCTTAACCTGTGGAAAAATGTACAGCAGACCCTGGGCCCTCATGTGACCTC	360
Oy	376	tgttcctccctccctgatatacataaacccaactcccttgttaccaattctctctgagctcat	435
Dd	361	TGTTCTCTCTCCCTCGATACATAAACCCCACTCTGTATTCAATTTCTCTCTGAGACTCAT	420
Oy	436	ggacacagatcttatgattatcatcatccctctgscaga	476
Dd	421	GGACACAGTATCTTGATTTATTCATCCCTGTGTACCAAGA	461
RESULT	3		
LOCUS	AL528423	905 bp	mRNA linear EST 13-FEB-2001
DEFINITION	AL528423 LTR_NF1003.NBC3 Homo sapiens CDNA clone CS0DC029YH22 5		
ACCESSION	AL528423		
VERSION	AL528423.1	GI:12791916	
KEYWORDS	ESF.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1. (bases 1 to 905)		
JOURNAL	L.I., M.B., Gruber, C., Jesssee, J. and Polayes, D.		
COMMENT	Full-length cDNA libraries and normalization unpublished (2001)		
FEATURES	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr. Location/Dualifiers		
Source	1..905		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CS0DC029YH22"		
	/clone_lib="LTR_NF1003.NBC3"		
	/sex="male"		
	/tissue_type="neuroblastoma cells"		
	/lab_host="DH10B"		
	/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifestech.com URL : http://fulllength.invitrogen.com"		
BASE COUNT	183 a 302 c 273 g 145 t 2 others		
ORIGIN			
Query Match	40.9% Score 449.6; DB 9; Length 905;		
Best Local Similarity	80.1% Pred. No. 9.1e-47;		
Matches 596; Conservative 1; Mismatches 75; Indels 72; Gaps 3;			
Oy	12	gagggccatgctgttggagctgcagacgsgttagtgcgaacacccgagctgcggaagaagaa	71
Dd	2	GCGAGCCATGGGGCTGGGGCTGCACGCGCAAGGTGACACGAGCTGCGCAAGGAAA	61
Oy	72	gtcgcggaacgcgcgccgcagccgcgcgcagccgcgaagaagcgaagtgctgtacacgtgc	131
Dd	62	GTCCCGGAGTGGCGGCCGACAGCCGSCGA--CAGAGACACAGAGTGTCTACACACTGCG	119
Oy	132	gcacactctgcccttgcgcgcgcgcgtcacgcgcgacacttgaaacagccctcatcatcg	191

Db	120	TCACACGCTGCCCTTTCGCCCCGGCGTACGGCCACCCTGGACAAAGCCTTATCATC	179
Qy	192	cctccacaatcagctactacctcgcattgcacgcgcctctgcg-----agcaggt	237
Db	180	CCTGACCATCAGACTACCTGCGGATGACACCGCCTTGCGCCGACAGGAGTGGAACACAGT	239
Qy	238	ggaaaaaggggagagagccacttgcgcgcctgtactactgaaggcccttgaagggttgc	297
Db	240	GGGAGCAGGGGGAGAACCATCTGGAGTGGCTGTACTGTGAAGCCCTGGAGGGCTTCGTAT	299
Qy	298	gttactaacccgcggaggagacatgctgtactctgcggaataatgttcaagaacaccttgcg	357
Db	300	GGTGTCTACCGCCGAGGAGACATGCGCTTACTGTGCGAGAAATGTACGAACACACTGGG	359
Qy	358	cctcagtcagtgcgaactctgtctcctcctctgaatacaccaccactcctgtgatacaat	417
Db	360	CCTGAGTGCAG-----	369
Qy	418	ttctctctgcgagctcatttgcagacagtaactcttgaattatccactcctgtgaccaagag	477
Db	370	-----CTGGAGGCTCATTTGGACACAGCATCTTTGATTTTCATCCACCCTGTGACCAAG	423
Qy	478	gaactcaagagcgcgccttaccccccagcgccgaacctgtataaagaagaacttgaagcccca	537
Db	424	GAGCTTCAGGACGCGCTTACACCCCCAGAGACCCCTGTCCAGAGAGGAAGTGGAGGCCCC	483
Qy	538	acaagagcgcacactcttccctgcgaaatgaagaagcagctccacagcagaagggcgacgctc	597
Db	484	ACGAGAGCGGTGCTTCTCTGCGATGAAGATACACTACACAGCGGGGGCGACACCTTC	543
Qy	598	aacctcaagcggccacacttgaaggtgtctgcactgtctcagaacatataagaagcctacaag	657
Db	544	AACCTCAAGGCGGCGCACCTGGAAAGTGTCTGAACCTCTCTGGACATATGAGGGCCCTCAAG	603
Qy	658	ccccctgcgaagacttccccctgcggaggccctcgcctcgcggcctccccctgaatgctctg	717
Db	604	CCACTGGCGAGACTTCTTCACGCTGGAGCCCTGTACTGAGAGCCCCCGCTCAGTGCCTG	663
Qy	718	gtgcttactcttgaagccatcccc	741
Db	664	GTGCTCATCTCGGAASCATCCCC	687
RESULT	4		
LOCUS	AL519496	856 bp	mRNA
DEFINITION	AL519496 LTI_NFL004_NBC2	Homo sapiens	CDNA clone CSDB004Y122
ACCESSION	AL519496		
KEYWORDS	AL519496.1	GI:12782989	
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1. (bases 1 to 856)		
JOURNAL	Li, W.B., Gruber, C., Jesses, J. and Polayes, D.		
COMMENT	Full-length cDNA libraries and normalization		
	Unpublished (2001)		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRY cedex - France		
	Email: segret@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
FEATURES			
Source	1..856		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CSDB004Y122"		
	/clone_lib="LTI_NFL004_NBC2"		
	/sex="male"		
	/tissue_type="neuroblastoma cells"		
	/lab_host="DH10B"		

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : filiang@lifetech.com URL : <http://fulllength.invitrogen.com>

BASE COUNT 169 a 289 c 257 g 139 t 2 others  
ORIGIN

Query Match 40.4%; Score 444.8; DB 9; Length 856;

Best Local Similarity 80.2%; Pred. No. 3.6e-46;

Matches 550; Conservative 1; Mismatches 73; Indels 72; Gaps 3;

```

QY 20 tggcgcttggtctgcagcgctgagtcgaacacgcagctgcggaagagagtcgcg99 79
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TGGCGCTGGGGCTGCAGCGCGCAAGGTCGACACGAGCTGCCCAAGAAAGGCCCGG 60

QY 80 acgagcgccgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 139
    | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ATGCGGCCCGCACCGCGCGCA--CAGGAGACCAAGGTGCTACACAGCTGCTACACGC 118

QY 140 tgcaccttgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 TGGCCTTGGCGCGCGGCGTACGCGCCACCTGGACAAGGCTCTATCATGCGCTACCA 178

QY 200 tcagctactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 245
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 TCAGCTACCTGCAGCATGACACGCGCTGCGCGCGAGGAGTGAGACAGTGGGAGCAG 238

QY 246 gggggaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 305
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 GGGGAGAACCACTGGAGCTGCTGCTGCTGAGGAGGCTGAGGAGGCTGCTGCTGCTCA 298

QY 306 ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 365
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 299 CCGCCGAGGAGACATGCTTACCTGTCGAGATGTCACCAAAACACCTGGGCTCAGTC 358

QY 366 agtgcacctgtctctctctctctctctctctctctctctctctctctctct 425
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 AG-----CT 362

QY 426 gggagctattggacacagtaattttatccatccctgtgaccaaagagaaattca 485
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 GGAAGCTCATTTGACACAGCATCTTTGATTTTCATCCACCTGTGACCAAGAGAGCTTCA 422

QY 486 agagcccttgaccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 545
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 GGAAGCCTTACCCCGACAGACACCTGTCCAGAGAGAGAGTGAAGGCCCGCCACAGGAG 482

QY 546 ccaatttccctgcgaatgaagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 605
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 483 GTGCTTCTCTTGGGATGAGAGATACACTACACAGCGGGGGGCGACCTCAACCTCAA 542

QY 606 agcgcgacacttgaaagtgtctgcacgtctcagagacatgaaggcctacaagcccttgc 665
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 GGGGGCCACCTGGAAGGTGCTGAACTGCTCTGGACATATGAGGGCTTACAAAGCCACTGC 602

QY 666 aacgaactcccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 725
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 603 GCAGACTTTCCTCCAGTGGAGCCCTTGACTCAGAGCCCGCGCTGAGAGCTGCTGCTCAT 662

QY 726 ctgtgaagcattcccc 741
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 663 CTGGAASCATCCCC 678

```

RESULT 5  
AL541807

LOCUS AL541807 919 bp mRNA linear EST 16-FEB-2001  
DEFINITION AL541807 LNI.PL002.Pl1 Homo sapiens cDNA clone CSDE007YA24 5 prime  
ACCESSION AL541807  
VERSION AL541807.1 GI:12873233  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 919)  
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).

FEATURES  
source  
1..919  
location/Qualifiers  
1..919  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSDE007YA24"  
/clone\_lib="LNI.PL002.Pl1"  
/lab\_host="DH10B"  
/note="Organ: Placenta; Vector: pCMVSPORT 6; 1st strand  
cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-stranded cDNA was digested with Not I  
and cloned into the Not I and Eco RV sites of the  
pCMVSPORT 6 vector. Library was constructed by Life  
Technologies. Contact : Feng Liang Life Technologies, a  
division of Invitrogen 9800 Medical Center Drive Rockville  
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :  
[filiang@lifetech.com](mailto:filiang@lifetech.com) URL :  
<http://fulllength.invitrogen.com>"

BASE COUNT 191 a 309 c 275 g 141 t 3 others  
ORIGIN

Query Match 38.8%; Score 427; DB 9; Length 919;

Best Local Similarity 78.9%; Pred. No. 5.5e-44;

Matches 576; Conservative 1; Mismatches 82; Indels 71; Gaps 3;

```

QY 26 tggggtgcagcgctgagtcgaacacgcagctgcggaagagagtcgcgcgcgcgc 85
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 35 TGGAGTGGCAAGACCAACAGGCTGACCAAGAGCTGCCAAGAAATCCCGGATCGG 94

QY 86 ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 145
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95 CCCGACCGCGCGCA-CAGNAGACCGAGAGTGTGCTGACAGCTGCTCACACGCTGCCCT 153

QY 146 ttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 205
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154 TCGCGCGCGGCTGACAGCGCCACCTGTCGACCAAGGCTCTATCATGCGCTCAGCT 213

QY 206 acctgcgatgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 214 ACTTGCACATGACACCGCTGTCGCGCGCAGGAGTGGAAACAGAGTGGAGCGAG 273

QY 252 agccactgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 AACCACTGATGCTGCTTACTGTAAGGCCCTGGAGGGCTTGCTATGCTCTCACGCCG 333

QY 312 agggagaacatgcttactctgcgaaatgtcaagcaacacatgcgcgcgcgcgcgcgc 371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 AGGGAGACATGGCTTACTCTGAGAGATGTCAAGAAACATGAGGCTCAGTCAG---- 389

QY 372 cctctgtctctctctctctctctctctctctctctctctctctctctctctga 431
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 390 -----CTGAGAGCT 397

QY 432 cattggaacagtatcttattatccatccctgtgaccaaagagaaattcaagcgc 491
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Query Match	37.1%	Score 408.2	DB 10	Length 767
Db	398	CATTGGACACGACATCTTTGATTTCATCATCCCGCTGTGACCAAGAGGAGCTTCAGAGACG	457	
QY	492	cctgaacccccgaagcccggaactctgtcaagaagaagctggaagccccaacagagcgcaact	551	
Db	458	CTTGACCCCCCGACGACAGACCTGTCCAGAGAGAAAGGTGGAGCCCCACAGGACGGTGTCTT	517	
QY	552	ttcccttggaatgaagagcagcgtccacagcagaagggcgcaagctcaactcaaaagcgac	611	
Db	518	CTCTCTGGCGATGAAGAGACGCTACCAAGCCGCGGCGCCACCTTCATCAAGCGGCG	577	
QY	612	caaccggaagtgtctgcgaactgtctcaggaacatatgaggcgctcaaacagccctgcgaacac	671	
Db	578	CACCTGGAAAGGTGCTGCACTGCTCTGGACATATTGAGGCGCTCAAGCCACCTGCGCAGAC	637	
QY	672	ttcccttcgaggaagcctcgcctcgcgaagcctcccttcgaatgcctgtgttatactgtga	731	
Db	638	TTCCTCAGATGGAGACCTGCACTCAGAGCCCCCGCTGCGATGCTCGTGTCTCATCTGCGA	697	
QY	732	agccatcccc 741		
Db	698	ASCATCCCC 707		
RESULT	6			
LOCUS	BG699633	767 bp	mRNA	linear
DEFINITION	6026793765P1 NIH_MGC_95 Homo sapiens CDNA clone IMAGE:4811824 5', mRNA sequence.			
ACCESSION	BG699633			
VERSION	BG699633.1	GI:13968143		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 767)			
TITLE	NIH-MGC htcp://mgc.nci.nih.gov/.			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bbs-rtm1.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki Toshiyuki and Piero Carninci (RIKEN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLM10703 row: b column: 17 High quality sequence stop: 616. Location/Qualifiers			
FEATURES	1..767			
SOURCE	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:4811824"			
	/clone_lib="NIH_MGC_95"			
	/tissue.type="hippocampus"			
	/lab_host="DH10B"			
	/note="Organ: Brain; Vector: pBluescript (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcagag ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3', size-selected for average insert size 2.5 kb and normalized to R0F 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC library."			
BASE COUNT	156 a	259 c	242 g	110 t

[illegible]



Db	118	ACTCTGCCCTTTGGCCGGGGCCCTCA - CACNTNCTGGACAAAGCCTCCATCATATGGCCTC	176
QY	196	acaatcaagtactatcgatgcagatgcacgcgcctctgcgcagcagbtygaaaaaaggysgaagcc	255
Db	177	ACAATCAGCTACTCTGCAGATG - ACCGCCTCTGGCAGACAGANTGAGAAAAAGGGGAGAGACC	235
QY	256	actgtagcgcgcgtactaccggaagccctctggaaggttttgtcatatgtaactacccgcggaag	315
Db	236	ACTGGACCCCTGCTACCTGAAGGCCCTGGAAGGTTTGCTCAAGTGATACACCGCCGAGAGG	295
QY	316	agacatgcttaacctgtctcgaaaatgctagcaagcaccctgagccctcagtcagtygaacct	375
Db	296	AGACATGCTTACCTGTGTGGAAAAATGTACGACGACCTGGGCTTCAGTCAATGGAACTTC	355
QY	376	tgcttcctccctctgatacataaaccccaactccctgttacaattctctcttgaagctcat	435
Db	356	TGTTCTCTCTCCCTCGATTCATATMACCCCACTCCTGTACCAATTTCTCTGTGAGACTAT	415
QY	436	ggaccacagatattcttgattatcatcatccctctgtgccaagggaaac	481
Db	416	GGACACAGATATCTTGATTTTAT - CATCCCTGTGACCAAGAGAGAAC	460

[illegible]

ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE	1 (bases 1 to 633)
	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE	Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D.
COMMENT	

Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Plate: LLM11523 row: g column: 23  
High quality sequence stop: 631.

FEATURES	Location/Qualifiers
source	1. .633

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5208622"
/clone_1fb="NH_MCC_122"
/lab_host="DH10B"
/notes="Organ: pooled lung and spleen; Vector: pCMV-Sport6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:

```

BASE COUNT	119 a	225 c	189 g	100 t
ORIGIN				

Query Match      30.5%;    Score 336;    DB 10;    Length 633;

Best Local Similarity	79.28;	Pred. No.	9.7e-33;
Matches	456;	Conservative	0;
		Mismatches	50;
		Indels	70;
		Gaps	2;

OY	44	gggaagaccgaagctcgaggaagaagatgctgggaagcgccgcgaagcggaac	103
Db	114	ggatgaccacgagctgcgtccgcaaggaaagttccgggattgcggccgcgacgacc	173
OY	104	aggaagcgaagtgctgctgaccagctgagcacaactctgaccttggcgcgagctlaagcg	163
Db	174	aggaagaccgaggtcgtgtaaccagctgacctcaacagctgccttggcccgcgacagc	233
OY	164	cgcacatggacaagcgctccatcatatgctgcgtccaaactaactgactcgtcgtatgcacgcgc	223
Db	234	cccacctcgacacaagccctctatctatgcccctccaccatagcttacctggcgtatgcacgcgc	293
OY	224	ctctgcgc-----agcagtggtgaaaaaggggagaggaacatgaacgctgct	269
Db	294	ttctggccgccgacagggagatggaaccaagggtggagcagggggagaaacactgattgcttctct	353
OY	270	acctgaagcgccctggaaggtcttcgtcaatgctactacccgcgaggaagacaatgactaac	329
Db	354	acctgaagcgccctggagggcttgcgtcattgagctcaccgcgagggagacatagctttacc	413
OY	330	tgtcgaaatgctcaagcaagcactgtagcctcaagtctgaatctgcttccctccct	389
Db	414	tgtcggaagaaatgtcagcaaacacactggggcccttcagtcag-----	451
OY	390	gataataaccacaactcctgtgtaccaattctctcttgtagctcatatggacaagatatt	449
Db	452	-----ctgagactcatttggacacacagactttt	477
OY	450	tgattttatcatccctgtgaccaaagagaacttcaagacgccttgacccccagcgca	509
Db	478	tgatatttattccaccctctgtgaccaaagagagcttcagagacggccttgacccccacagcagac	537
OY	510	ccctgtcaagaagaagcttggaaagccccaacagacgctccacttctccctgcgaatgaagag	569
Db	538	cctgtccgaggaagaaaggtgtgagggcccccacgacgaggggtcttcttgcgcatgaaagag	597
OY	570	caagctacagcagagggcgcaagcttaactcaa	605
Db	598	ttacgctcacacagccggggccgacaccttaacctcaa	633

RESULT 10

AA041707	AA041707	459 bp	mRNA	linear	EST 03-SEP-1996
LOCUS					
DEFINITION	mj02g02.t1 Soares mouse embryo NME13.5 14.5 Mus musculus cDNA clone IMAGE:474962 5', mRNA sequence.				

ACCESSION	AA041707
VERSION	AA041707.1
KEYWORDS	GI:1519863
SOURCE	EST.
	house mouse.

ORGANISM	Mus musculus
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:	
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.	
1 (bases 1 to 459)	
REFERENCE	
AUTHORS	Matra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

TITLE	JOURNAL	COMMENT
The WashU-HMMI Mouse EST Project	Unpublished (1996)	Contact: Marra M/Mouse
Gelsel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, B., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.		

Washington-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
This clone is available royalty-free through LLNT; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.





QY	1034	cag	1036	
Db	3	CAG	1	
RESULT 12				
LOCUS	BM118909/c			
DEFINITION	BM118909	425 bp	mRNA	linear
ACCESSION	BM118909			EST 01-FEB-2002
VERSION	BM118909.1			
KEYWORDS	musculus			
SOURCE	house mouse.			
ORGANISM	house mouse.			
REFERENCE	Eukaryote: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 425)			
AUTHORS	Piao, Y., Karcul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K. and K.O.M.S.H.			
TITLE	Systematic Analyses of NIA Mouse Newborn Kidney cDNA Library			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Dawood B. Dudekula Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: cdna@igsun.grc.nia.nih.gov Plate: L0917 row: H column: 06 Seq primer: -21M13 Forward High quality sequence stop: 425 POLYA-Yes.			
FEATURES	Location/Qualifiers			
SOURCE	1. 425 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="niaEST:L0917H06-3" /db_xref="taxon:10090" /clone="L0917H06" /clone_lib="NIA Mouse Newborn Kidney cDNA Library (Long)" /issue_type="Newborn Kidney" /dev_stage="Newborn" /lab_host="DH10B" /note="Organ: kidney; Vector: pSPORT1 (Invitrogen); Site.1: Salt; Site.2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11:1553-1558 (2001). [PMID:11544199]). In brief ,double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5'-pGACTAATTCAGATCCGACGCCGCCCTTTTGTGTTT-3'] from 26 microgram of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-salt4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Salt4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with Salt and NotI enzymes, and cloned into Salt/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with ligation mixture by the standard chemical method. The average insert size is about 3.0 kb. The library was constructed by Yulan Piao (NIA)."			
BASE COUNT	95 a	96 c	136 g	98 t
ORIGIN				
Query Match	27.1%	Score 298.2;	DB 10;	Length 425;
Best Local Similarity	99.0%;	Prod. No. 4.8e-28;		
Matches 300;	Conservative	0;	Mismatches 3;	Indels 0;
			Gaps	0;

QY	734	ccatccccagctcccccctcccaagatggtgctaactcttggtctcccaaggagaactc	733
Db	303	TTAACCCCCAGAGTCCCTTTCCACGATGGTGTCACTGGGTCTTCCACAGGAGAACTC	244
QY	794	ccatctcaccctatcacccctctcttggaaggcactacttgctgtccaagggtgc	853
Db	243	CCATCTCTACCTTATTACCCCTCTTTGGAGGCACTACTTTGTCCTTGCAAGAGGTGC	184
QY	854	ctgtccaggtgtctacaagggaaggagactgtacctctctccctcatgggtgtgc	913
Db	183	CTGTTCAGTGTCTACAGGGGAAAGGAGCTGAATCTCTCTCCCTCATGGGTGTGGG	124
QY	914	cccttaaccggaaaaattgtctctggttaaggaggaatgtgaagacatggcccaactctc	973
Db	123	CCCTTTAACCGGAAAAATTTCTCGGTACGAGGAGGAGGACATGAGCCACTTACTCT	64
QY	974	tagcccaagaaccacaaatgtctcccaaaaccacataaagacctctcctgttaggcac	1033
Db	63	TAGCCCAAGAAACCCACAATGTCTCCAAAACCACTAAAGACTCTCTCTGTTAGGCAC	4
QY	1034	cag 1036	
Db	3	CAG 1	
RESULT	13		
LOCUS	BM119659	610 bp	linear
DEFINITION	L0929A01-3 NIA Mouse Newborn Kidney cDNA Library (Long)	Mus	
ACCESSION	BM119659		
VERSION	BM119659.1	GI:17087685	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 610) Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K. and Ko, M.S.H.		
TITLE	Systematic Analyses of NIA Mouse Newborn Kidney cDNA Library		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Dawood B. Dudekula Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: cdna@igsun.grc.nia.nih.gov Plate: L0929 row: A column: 01 Seq primer: -21M13 Forward High quality sequence stop: 610 POLA=yes.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..610		
	/organism="Mus musculus"		
	/strain="C57BL/6J"		
	/db_xref="taxon:10090"		
	/clone="L0929A01"		
	/clone_11b="NIA Mouse Newborn Kidney cDNA Library (Long)"		
	/tissue_type="Newborn Kidney"		
	/dev_stage="Newborn"		
	/lab_host="DH10B"		
	/note="Organ: Kidney; Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA) Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11:1553-1558 (2001). [PMD:11544199]). In brief ,double-stranded cDNAs were synthesized with an Oligo(dt) primer (Invitrogen: 5'-		
	gcgcagcttctagatgcgagcgccgccttttttttttt-3') from 26		



TITLE HRI human cDNA project  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp

HRI human cDNA project; 5'-6 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

FEATURES  
source location/Qualifiers

1..873  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEMBA1000216"  
/clone\_lib="HEMBA1"  
/tissue\_type="whole embryo, mainly head"  
/dev\_stage="embryo, 10 weeks"  
/note="Vector: pME18SFL3"

BASE COUNT 175 a 305 c 248 g 142 t 3 others  
ORIGIN

Query Match 23.0%; Score 253.2; DB 9; Length 873;  
Best Local Similarity 86.6%; Pred. No. 1.5e-22;

Matches 279; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 424 ctgagagctaatggaacagatatttgaattatccatccctgtgaccaaagaagaatt 483  
DB 281 ctggagctcatgtgacacagcatcttgatttcattccaccctgtgaccaaagaagctt 340  
OY 484 caagaagccctgaaccccaagccgaacctgtcaaaagaagaagctggaagccccaacagag 543  
DB 341 caggacgcccctgagcccccagacacctgtccagagagagagtgagagcccccacagag 400  
OY 544 cggcactttccctcgtgaatgaagaagcaccacagcagaagggcgcaagctcaactc 603  
DB 401 cggctcttctccttgccgcatgaagagtacgtaccacgcccggcgccacccctcaactc 460  
OY 604 aaagcggccacctggaaggtgtcgtcagctcaggaacatagaaggcctacaagcccccct 663  
DB 461 aagcggccacctggaaggtgtcgtcagctcaggaacatagaaggcctacaagcccccct 520  
OY 664 gcaacagacttccctcgtcggcggagacccctcgtcgaagcctccctgcaatgctgtgctt 723  
DB 521 gccgagactttctccagctggagccctgactcagaagcccccctgctgagtgctgtctc 580  
OY 724 atctgtgaagcattccccagc 745  
DB 581 atctgtgaagcattccccagc 602

Search completed: September 8, 2002, 05:57:29  
Job time: 6842 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 05:34:17 : Search time 2212.99 Seconds  
(without alignments)  
10401.850 Million cell updates/sec

Title: US-09-896-791B-2  
Perfect score: 1100  
Sequence: 1 gaattcgccagcgagccat.....aaaaaaacatgcgccgc 1100

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0', Gapext 1.0

Searched: 1797656 segs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1032	93.8	1032	10	AF416641
2	556.6	50.6	2207	10	AF060194
3	505.4	45.9	2077	10	RN0277827
4	471.6	42.9	2082	9	AK027725
5	469.4	42.7	2023	9	AF463492
6	451	41.0	2004	9	AB054067
7	447.4	40.7	2595	9	AK021653
8	301.6	27.4	250615	2	AC073814
9	276	25.1	134189	2	AC079550
10	265.4	24.1	345	9	AF079154
11	253.4	23.0	1708	9	AK021421
12	228.2	20.7	167150	2	AC095259
13	211.6	19.2	1841	9	AK024095
14	196	17.8	134189	2	AC079550
15	192	17.5	3605	5	AF304864
16	191	17.4	191	10	MMHIF3A02
17	176.8	16.1	2888	4	AB018399
18	174.6	15.9	2730	9	HS051626
19	173	15.7	2816	6	184559
20	173	15.7	2818	9	HS081984
21	173	15.7	146180	9	AC007193
22	172	15.6	2776	6	AX330581
23	169.6	15.4	3488	10	D89787
24	168	15.3	3415	10	AF045160
25	160	14.5	2730	10	RN0277828
26	156.6	14.2	3031	6	184560
27	156.6	14.2	3033	10	MM081983
28	148.8	13.5	6597	5	AB013746
29	146.2	13.3	3718	10	AF057308
30	144.6	13.1	2711	10	RNHYPFAC1
31	139.8	12.7	3973	10	AF003695
32	138.2	12.6	3746	6	AX306008
33	138.2	12.6	3746	10	MMHIF1ALP
34	136.4	12.4	2640	5	AF129813
35	136.2	12.4	2551	4	AB018398
36	135.8	12.3	146	10	MMHIF3A03
37	135.4	12.3	3076	5	AF212989
38	135	12.3	3867	10	MM059496
39	134.6	12.2	2509	9	AF304431
40	134.6	12.2	2522	9	AF208487
41	134.6	12.2	2528	6	AX230580
42	134.6	12.2	2530	9	AF207601
43	134.6	12.2	2537	9	AF207602
44	134.6	12.2	3551	9	AB073325
45	134.6	12.2	3678	9	HSU22431

## ALIGNMENTS

RESULT	1	AF416641	1032 bp	MRNA	linear	ROD 09-DEC-2001
LOCUS	AF416641					
DEFINITION	Mus musculus inhibitory PAS domain protein (Ipas)					complete
ACCESSION	AF416641					
VERSION	AF416641.1	GI:17432318				
KEYWORDS						
SOURCE	house mouse.					
ORGANISM	Mus musculus					
REFERENCE						
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
TITLE	1 (bases 1 to 1032)					
JOURNAL	Makino, Y., Cao, R., Svensson, K., Bertilsson, G., Asman, M., Tanaka, H.,					
PUBMED	Caoy, Y., Berkenstam, A., and Poellinger, L.					
REFERENCE	Inhibitory PAS domain protein is a negative regulator of					
AUTHORS	hypoxia-inducible gene expression					
TITLE	Nature 414 (6863), 550-554 (2001)					
Direct Submission						



BASE COUNT 471 a 742 c 597 g 397 t  
ORIGIN





Db	74	GGAAAGATGCCGGATGCGGGCCCGACGCCGGCGGACAGGACCGAGTGCTTACCA	133
OY	126	gctgycacacactctgcctcttgcgcgcgcgtcagcgcgcacttgcgaagaagcctcat	185
Db	134	GCTACTCACAAGCGTGGCCCTTGCCCGGGGTGAGGCCACCGTGGAAAGGCGCTTAT	193
OY	186	catgcgcctcaaatcagctactcgtgcgtatgcacgcgcctctgcgc-----a	231
Db	194	CATGGCGCTTCACATACACTACTCTCCGATCCACCGCTTCGCGCGGAGGAGTGGAA	253
OY	232	gcaggttgaaaaaaggggagagacacacttgcgcgcgtcactcttgaagggccttgcgagt	291
Db	254	CGAGTGGGAGAGGGGGGAGAAACCTGGATGCCCTGCTACCTGGAAGGCGCTGGAGGCTT	313
OY	292	cgtcatgtactcacccgcgcgaggagacatgcttactgttcgcgaanaatgtcagaacga	351
Db	314	CGTATGGTGGCTCACCGCGGAGGAGACATGGCTTACTGTGCGGAGATGTACGACAAACA	373
OY	352	cctgagcctcagtcagtygacacctgtcttcctcctccctgcataacataaccacccactcgt	411
Db	374	CCTGGGCTTCAGTCG-----	389
OY	412	accaatctctctcagagctcatgtgacacagatccttgaatttcatccctctgtac	471
Db	390	-----CTGGACTCATTTGGACACAGCATCTTCGATTTTATTCACCCCGTGTAC	437
OY	472	caagaagaacttcaagaagccttgaccccccaagcgcgaactgttcaagaagaagaagcttgaa	531
Db	438	CAAGAGGAGCTTTCAGAGAGCGCCCTGTACCCCGCAGCAGACCCCTGTCCAGGAGGAAGGTGAG	497
OY	532	gcccaacaagcgcgaacttctcctgtggaatgaagaacacgctcaacgaagagagcgc	591
Db	498	GCCCCACAGGAGCGGTGCTTCTCTCTTTGGCAGATGAAGAGATTAACACACAGCGCGGGGCGC	557

Oy	592	acgtcaaccctcaaaagcgccacccggaagtgtgcgaactctcaaggacatattggggc	651
Dd	558	ACCCTCAACCTTCAGGCGGCCACACTGGAAAGTGCTGAACCTGCTTGTGGACATATTGAGGGCC	617
Oy	652	tacaagccccctgtcacagaacttcaccttgcgcggagacccctgcctccgaagctcccctgcaa	711
Dd	618	TACAAGCACACCTGCAGCAGACTTCTCCAGCTGGAGACCCCTGACTCAGAGCCCCCGCTGAG	677
Oy	712	tgcttgtgtctatctgtgaagccaatcccccaagctcccttcacaga	758
Dd	678	TGCCGTGTCATCTCATCTGCCAAGCCATCCCACCACGAGCAGCCTTGA	724
RESULT	6		
AB054067			
LOCUS	AB054067	2004 bp	mRNA linear PRI 18-JAN-2002
DEFINITION	Homo sapiens HIF-3A mRNA for hypoxia-inducible factor-3 alpha,		
	complete cds.		
ACCESSION	AB054067		
VERSION	AB054067.1	GI:16040974	
KEYWORDS			
SOURCE	Homo sapiens kidney cDNA to mRNA.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	Hara,S., Hamada,J., Kobayashi,C., Kondo,Y. and Imura,N.		
	1 (sites)		

JOURNAL MEDLINE  
2145827  
expression by HIF-1 $\alpha$   
Biochem. Biophys. Res. Commun. 287 (4), 808-813 (2001)  
2 (bases 1 to 2004)  
Hara, S., Kobayashi, C., Hamada, J., and Imura, N.  
Direct Submission  
Submitted (15-JAN-2001) Shuntaro Hara, Kitasato University, School  
of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo  
106-8641, Japan (E-mail: haras@pharm.kitasato-u.ac.jp,  
Tel.:81-3-5791-6265, Fax:81-3-3442-4146)







COMMENT  
(E-mail:genomics@ri.co.jp, Tel:81-338-52-3951, Fax:81-338-52-3952)  
NEO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan: cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5' - & 3'-end one pass sequencing and clone selection:  
Helix Research Institute (supported by Japan Key Technology Center  
etc.) and Department of Virology, Institute of Medical Science,

```

FEATURES             University of Tokyo.
Source               Location/Qualifiers
                    1..1708
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="HEMBA1000216"
                    /tissue_type="whole embryo, mainly head"
                    /clone_lib="HEMBA1"
                    /dev_stage="embryo, 10 weeks"
                    /note="cloning vector: pME18SFL3"
                    125..724
CDS
                    /note="unnamed protein product"
                    /codon_start=1
                    /protein_id="BAB13819.1"
                    /db_xref="GI:10432601"
                    /translation="MRPAGAARBRCTSMTRCPSAPASPTWRPLSCASPSATC
                    ACTASPOLIELGHSIRDFHPCDEEODALTPOOTLSRKVEAPTRCSTLRKST
                    LTRSRRTNLKATKYNKVCNGSMHAKYRPPAOTSFPAGSPDEPPLQCLVLCALPFP
                    GSLEPPLGRGAFSLSHSDMKFTYCDNVAEAVAGSPMT"
                    280..281
misc_feature
                    /note="146 bases segment is present in AK21653, probably
                    caused by alternative splicing."
misc_difference
                    /note="compared to AK21653 and AC007193.1"
                    /replace="c"
BASE COUNT          340 a      588 c      461 g      319 t
ORIGIN

Query Match          23.0%; Score 253.4; DB 9; Length 1708;
Best Local Similarity 84.8%; Pred. No. 2.6e-41;
Matches 284; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Oy 424 ctggagcatttgagacagatatttattatccatccctgtaccagaagaatt 483
    |||||||
Db 281 ctggagcatttgagacagatatttattatccatccctgtaccagaagaatt 340

Oy 484 caaagcgccttgaccccgagcgaactgtcaaaagaagctgaaagcccaacagag 543
    |||||||
Db 341 CAGGAGGCCCTGACCCCGCAGAGACCCCTGTCCAGAGAGAGAGAGAGAGAGAGAG 400

Oy 544 cgcacatttccctcggaatgaagagcagctccaccagagagggcgacgtcaactc 603
    |||||||
Db 401 cgcctcttctcttgcgcatgaagatgacgtccaccagcggcgccacctcaactc 460

Oy 604 aaagcgcgcactggaaggtgtcgcagctgcagagacatagaggccctacaagccctc 663
    |||||||
Db 461 AAGGCGGCCACTGGAAGGTGCTGTGACATATGAGGGCCCTACAAAGCCACT 520

Oy 664 ggcagacattccctcgagcggagccctcgctccgagcctccctcgcaatgctgtgctt 723
    |||||||
Db 521 GCGCAGACTTTCAGCTGCGAGAGCCCTGACCTGACAGAGCCCGCGCTGAGTGTGCTC 580

Oy 724 atctgtgaagcattcccccagctcccttcacaga 758
    |||||||
Db 581 ATCTGCGAAGCCATCCCGCAGGCAAGCCCTGGA 615

RESULT 12
AC095259          167150 bp      DNA      linear      HTG 20-DEC-2001
DEFINITION       Rattus norvegicus clone CH230-10B18, *** SEQUENCING IN PROGRESS
ACCESSION        AC095259
VERSION          AC095259.2 GI:17943467
KEYWORDS         HTG; HTGS_PHASE1.
SOURCE           Norway rat.
ORGANISM         Rattus norvegicus
                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                 Rattus.
REFERENCE        1 (bases 1 to 167150)
AUTHORS          Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

```

```

TITLE            JOURNAL
JOURNAL          JOURNAL
AUTHORS          Wootley,K.C.
TITLE            Direct Submission
JOURNAL          Direct Submission
COMMENT          2 (bases 1 to 167150)
                  Wootley,K.C.
                  Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
                  of Molecular and Human Genetics, Baylor College of Medicine, One
                  Baylor Plaza, Houston, TX 77030, USA
                  On Dec 20, 2001 this sequence version replaced gi:15625813.
                  ----- Genome Center
                  Center: Baylor College of Medicine
                  Center code: BCM
                  Web site: http://www.hgsc.bcm.tmc.edu/
                  Contact: hgsc-help@bcm.tmc.edu
                  ----- Project Information
                  Center project name: GDX
                  Center clone name: CH230-10B18
                  ----- Summary Statistics
                  Assembly program: Phrap; version 0.990329first call to
                  findphraplist
                  Consensus quality: 128712 bases at least Q40
                  Consensus quality: 139536 bases at least Q30
                  Consensus quality: 148385 bases at least Q20
                  Estimated insert size: 136187; sum-of-contigs estimation
                  Quality coverage: 0x in Q20 bases; agarose-gel estimation
                  Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation
                  -----
                  * NOTE: Estimated insert size may differ from sequence length
                  * (see http://www.hgsc.bcm.tmc.edu/docs/genbank-draft-data.html).
                  * NOTE: This is a 'working draft' sequence. It currently
                  * consists of 67 contigs. The true order of the pieces
                  * is not known and their order in this sequence record is
                  * arbitrary. Gaps between the contigs are represented as
                  * runs of N, but the exact sizes of the gaps are unknown.
                  * This record will be updated with the finished sequence

```

\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 5997: contig of 5997 bp in length  
\* 5998  
\* 6098 6097: gap of unknown length  
\* 12971 12971: contig of 6874 bp in length  
\* 13072 13071: gap of unknown length  
\* 15335 15335: contig of 2264 bp in length  
\* 15336 15335: gap of unknown length  
\* 15436 21889: contig of 6454 bp in length  
\* 21890 21889: gap of unknown length  
\* 26955 26955: contig of 4966 bp in length  
\* 27056 27055: gap of unknown length  
\* 31738 31737: contig of 4682 bp in length  
\* 31838 31837: gap of unknown length  
\* 35634 35733: contig of 3796 bp in length  
\* 35734 35733: gap of unknown length  
\* 38571 38670: contig of 2837 bp in length  
\* 38671 38670: gap of unknown length  
\* 43208 43207: contig of 4537 bp in length  
\* 43308 43307: gap of unknown length  
\* 48004 48003: contig of 4696 bp in length  
\* 48104 48103: gap of unknown length  
\* 50815 50914: contig of 2711 bp in length  
\* 50915 50914: gap of unknown length  
\* 54186 54185: contig of 3271 bp in length  
\* 54286 54285: gap of unknown length  
\* 58647 58746: contig of 4361 bp in length  
\* 58747 58746: gap of unknown length  
\* 61144 61143: contig of 2397 bp in length  
\* 61244 61243: gap of unknown length  
\* 64597 64596: contig of 3353 bp in length  
\* 64697 67316: gap of unknown length  
\* 67317 67316: contig of 2620 bp in length  
\* 67417 67416: gap of unknown length  
\* 70771 70770: contig of 3354 bp in length  
\* 70871 70870: gap of unknown length  
\* 73845 73944: contig of 2974 bp in length  
\* 73945 73944: gap of unknown length  
\* 77641 77641: contig of 3697 bp in length  
\* 77742 77741: gap of unknown length  
\* 79929 79928: contig of 2187 bp in length  
\* 80029 80028: gap of unknown length  
\* 82787 82786: contig of 2758 bp in length  
\* 82887 82886: gap of unknown length  
\* 84591 84591: contig of 1705 bp in length  
\* 84592 84591: gap of unknown length  
\* 84692 87672: contig of 2981 bp in length  
\* 87673 87672: gap of unknown length  
\* 87773 90407: contig of 2635 bp in length  
\* 90408 90507: gap of unknown length  
\* 90508 93319: contig of 2812 bp in length  
\* 93320 93319: gap of unknown length  
\* 93420 96270: contig of 2851 bp in length  
\* 96271 96370: gap of unknown length  
\* 96371 98956: contig of 2586 bp in length  
\* 98957 98956: gap of unknown length  
\* 99057 100605: contig of 1549 bp in length  
\* 100606 100705: gap of unknown length  
\* 100706 104118: contig of 3413 bp in length  
\* 104119 104218: gap of unknown length  
\* 104219 105843: contig of 1625 bp in length  
\* 105844 105943: gap of unknown length  
\* 105944 108705: contig of 2762 bp in length  
\* 108706 108805: gap of unknown length  
\* 108806 110740: contig of 1935 bp in length  
\* 110741 110840: gap of unknown length  
\* 110841 112762: contig of 1922 bp in length  
\* 112763 112862: gap of unknown length  
\* 112863 115666: contig of 2804 bp in length  
\* 115667 115766: gap of unknown length  
\* 115767 117269: contig of 1503 bp in length  
\* 117270 117369: gap of unknown length  
\* 120158: contig of 2789 bp in length

\* 120159 120258: gap of unknown length  
\* 120259 122019: contig of 1761 bp in length  
\* 122020 122119: gap of unknown length  
\* 122120 124117: contig of 1998 bp in length  
\* 124118 124217: gap of unknown length  
\* 124218 126395: contig of 2178 bp in length  
\* 126396 126495: gap of unknown length  
\* 126496 127499: contig of 1004 bp in length  
\* 127500 127599: gap of unknown length  
\* 127600 129051: contig of 1452 bp in length  
\* 129052 129151: gap of unknown length  
\* 129152 130882: contig of 1731 bp in length  
\* 130883 130983: gap of unknown length  
\* 130983 132045: contig of 1063 bp in length  
\* 132046 132145: gap of unknown length  
\* 132146 133485: contig of 1340 bp in length  
\* 133486 133585: gap of unknown length  
\* 133586 135588: contig of 2003 bp in length  
\* 135589 135688: gap of unknown length  
\* 135689 137050: contig of 1362 bp in length  
\* 137051 137150: gap of unknown length  
\* 137151 138633: contig of 1483 bp in length  
\* 138634 138733: gap of unknown length  
\* 138734 140308: contig of 1575 bp in length  
\* 140309 140408: gap of unknown length  
\* 140409 141706: contig of 1298 bp in length  
\* 141707 141806: gap of unknown length  
\* 141807 143314: contig of 1508 bp in length  
\* 143315 143414: gap of unknown length  
\* 143415 144796: contig of 1382 bp in length  
\* 144797 144897: gap of unknown length  
\* 144897 146087: contig of 1191 bp in length  
\* 146088 146187: gap of unknown length  
\* 146188 147644: contig of 1457 bp in length  
\* 147645 147744: gap of unknown length  
\* 147745 149005: contig of 1261 bp in length  
\* 149006 149105: gap of unknown length  
\* 149106 150998: contig of 1893 bp in length  
\* 150999 151098: gap of unknown length  
\* 151099 152445: contig of 1347 bp in length  
\* 152446 152545: gap of unknown length

Query Match 20.7%: Score 228.2; DB.2: Length 167150;  
Best Local Similarity 87.9%: Pred. No. 6.8e-36;  
Matches 284; Conservative 0; Mismatches 33; Indels 6; Gaps 3;

QY 724 atctgtgaagcatcccccagctcccttccacgagatggtgtactctgtgtctccacag 783  
|| || |  
Db 59881 ATGCGTCTTCTAAACCCCAAGTCCCTTCTACATGATGCTACGCGGTCTCCACAG 59940  
|| || |  
QY 784 gagaagaactccatcttacttacttaccctcttgaaggact----acttgtct 839  
|| || |  
Db 59941 GA-AAGGCTCCCATCTCACTTATTCACCCCTCTTGGAAGCAATTCATATGTCCT 59999  
|| || |  
QY 840 tgtcaagaaggtgagctgtgttcagtgtcctacaggaaggaaggaatctctctccctc 899  
|| || |  
Db 60000 TGTCAAGAGGTGCTTGTTCAGGTCTTACAGGGGAAAGGCGCTGAATCTCTCCCTC 60059  
|| || |  
QY 900 atggtgtgtgtgagccttaaccggaataatgtctct-ggctaggagggagtgaagaca 958  
|| || |  
Db 60060 ATGATGTTATATAGCCCTTAATGGAAGAAATGTCTCGGTGAGGAGGAAGGAGACC 60119  
|| || |  
QY 959 tggccagctatctcttaagcagaaacacaaatgtctcctcaaaacacacataaaggact 1018  
|| || |  
Db 60120 TGGCCACAGTGTCTTACTCTCGAAACCAAAATGTCTCAAACTACCAATGAAGACTT 60179  
|| || |  
QY 1019 ctcctgtgtgagcagcagagaaa 1041  
|| || |  
Db 60180 CGCCTTGGTAGCACACAGAGACA 60202  
|| || |

RESULT 13  
AK024095

LOCUS AK024095 1841 bp mRNA linear PRI 29-SEP-2000  
 DEFINITION Homo sapiens cDNA FLJ14033 fis, clone HEMBA1004479, weakly similar  
 to HYPOXIA-INDUCIBLE FACTOR 1 ALPHA.  
 ACCESSION AK024095  
 VERSION AK024095.1 GI:10436389  
 KEYWORDS oligo capping, fis (full insert sequence).  
 SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to  
 mRNA, clone\_1lb:HEMB1 clone:HEMB1004479.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
 Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,  
 Nagatsuna,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,  
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,  
 Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,  
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,  
 Nakamura,Y., Nagahari,K., Masuho,Y., Nishimura,K. and Iwayanagi,T.  
 NEDO human cDNA sequencing project  
 TITLE Unpublished (2000)  
 JOURNAL 2 (bases 1 to 1841)  
 REFERENCE Isogai,T. and Otsuki,T.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)  
 NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing: Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing and clone selection:  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..1841  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HEMB1004479"  
 /tissue="whole embryo, mainly head"  
 /clone\_1lb="HEMB1"  
 /dev\_stage="embryo, 10 weeks"  
 /note="cloning vector: pME18SFL3"  
 119..832  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB14824.1"  
 /db\_xref="GI:10436390"  
 /translation="MRPAGAAARRPCTSWLTRCPSPAPASAPFTWRPLSCASPSATC  
 ACTASAPQELDALPQOTLSRRKVEAPTERFSLRMKSTLRSRTLNKATWV  
 LACTSGHRAVKKPAPSPAPSPDSEPLQCLVLCALPHPGSLRSEFVNLPRGDMV  
 YLSEVSNHLSGSOLELIGHSTFDFIHPDQDELQDALPLHLNTSSLLPKPGQTVSEFL  
 ABSYPRRSFHLPRWMP"  
 CDS  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB14824.1"  
 /db\_xref="GI:10436390"  
 /translation="MRPAGAAARRPCTSWLTRCPSPAPASAPFTWRPLSCASPSATC  
 ACTASAPQELDALPQOTLSRRKVEAPTERFSLRMKSTLRSRTLNKATWV  
 LACTSGHRAVKKPAPSPAPSPDSEPLQCLVLCALPHPGSLRSEFVNLPRGDMV  
 YLSEVSNHLSGSOLELIGHSTFDFIHPDQDELQDALPLHLNTSSLLPKPGQTVSEFL  
 ABSYPRRSFHLPRWMP"  
 BASE COUNT 404 a 514 c 519 g 404 t  
 ORIGIN  
 Query Match 19.2%; Score 211.6; DB 9; Length 1841;  
 Best local Similarity 83.1%; Pred. No. 7.5e-33;  
 Matches 241; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 QY 472 caaaggaactcaagcgcctgaccccaagcgcgaacctgtcaagaagaagctgaa 531  
 ||||||| ||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 272 CAAGAGAGCTTCAGACGCGCTGACCCCGACAGACCTGTGCCAGAGAGAGTGAG 331  
 QY 532 gccccaacagagcgcaactttccctcggaatgaagaagcagctaccagcagaggcgc 591  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 332 GCCCCACGAGGAGGGCTTCCTTCCTGCGCATGAAGAGTACGCTCACACCGCGGCGC 391  
 QY 592 agcgtcaacctcaagcgcagctcgaaggtgctgacgtccagacacatagaagcc 651  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 392 ACCCTCAACTCAAGCGCGCCACCTGGAAGGTGCTGAACATGCTGTGACATATGAGGCC 451

QY 652 tacaagccctgacagacttccctgacgagccctgctccagcctccctgcaa 711  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 452 TACAGCCACTGGCGAGACTTTCAGCTGGAGGCGCCCTGACAGAGCCCGCTCAG 511  
 QY 712 tgctgtgtctatctgtgaagcattcccccagctcccttcacagatgg 761  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 512 TGCTGTGTCTCATCTCGAAGCAGCATCCCGACCGAGAGCCTGAGGG 561  
 RESULT 14  
 AC079550/c 134189 bp DNA linear HTG 02-SEP-2000  
 LOCUS AC079550/c  
 DEFINITION Mus musculus clone RP23-422A19, WORKING DRAFT SEQUENCE, 30  
 unordered pieces.  
 ACCESSION AC079550  
 VERSION AC079550.1 GI:9964915  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DNAFT.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 134189)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Mouse  
 JOURNAL Unpublished  
 2 (bases 1 to 134189)  
 REFERENCE DOE Joint Genome Institute.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov  
 -----  
 Project Information  
 Center Project Name: 1898939  
 Center clone name: RPI-23\_422A19  
 -----  
 Summary Statistics  
 Consensus quality: 105629 bases at least Q40  
 Consensus quality: 114783 bases at least Q30  
 Consensus quality: 117545 bases at least Q20  
 Estimated insert size: 208000; agarose-fp estimation  
 Estimated insert size: 131289; sum-of-confits estimation  
 Quality coverage: 8.71 in Q20 bases; agarose-fp estimation  
 Quality coverage: 13.79 in Q20 bases; sum-of-confits estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 30 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1  
 \* 1340 1339: contig of 1339 bp in length  
 \* 1439: gap of unknown length  
 \* 1440 2741: contig of 1302 bp in length  
 \* 2742 2841: gap of unknown length  
 \* 2842 4408: contig of 1567 bp in length  
 \* 4409 4508: gap of unknown length  
 \* 4509 5699: contig of 1191 bp in length  
 \* 5700 5799: gap of unknown length  
 \* 5800 7306: contig of 1507 bp in length  
 \* 7307 7406: gap of unknown length  
 \* 7407 8863: contig of 1457 bp in length  
 \* 8864 8963: gap of unknown length  
 \* 8964 10169: contig of 1206 bp in length  
 \* 10170 10269: gap of unknown length  
 \* 10270 11579: contig of 1310 bp in length  
 \* 11580 11679: gap of unknown length  
 \* 11680 12992: contig of 1313 bp in length  
 \* 12993 13092: gap of unknown length



*	13093	14678:	contig of 1586 bp in length
*	14679	14778:	gap of unknown length
*	14779	17184:	contig of 2406 bp in length
*	17185	17284:	gap of unknown length
*	17285	19400:	contig of 2116 bp in length
*	19401	19500:	gap of unknown length
*	19501	21329:	contig of 1829 bp in length
*	21330	21429:	gap of unknown length
*	21430	24372:	contig of 2943 bp in length
*	24373	24472:	gap of unknown length
*	24473	25726:	contig of 1254 bp in length
*	25727	25826:	gap of unknown length
*	25827	28260:	contig of 2434 bp in length
*	28261	28360:	gap of unknown length
*	28361	29974:	contig of 1614 bp in length
*	29975	30074:	gap of unknown length
*	30075	34625:	contig of 4551 bp in length
*	34626	34725:	gap of unknown length
*	34726	36694:	contig of 1969 bp in length
*	36695	36794:	gap of unknown length
*	36795	41067:	contig of 4273 bp in length
*	41068	41167:	gap of unknown length
*	41168	42418:	contig of 1251 bp in length
*	42419	42518:	gap of unknown length
*	42519	47037:	contig of 4519 bp in length
*	47038	47137:	gap of unknown length
*	47138	55776:	contig of 8639 bp in length
*	55777	55876:	gap of unknown length
*	55877	62540:	contig of 6664 bp in length
*	62541	62640:	gap of unknown length
*	62641	69499:	contig of 6859 bp in length
*	69500	69599:	gap of unknown length
*	69600	77648:	contig of 8049 bp in length
*	77649	77748:	gap of unknown length
*	77749	86556:	contig of 8808 bp in length
*	86557	86656:	gap of unknown length
*	100999	100998:	contig of 14342 bp in length
*	101099	115677:	gap of unknown length
*	115678	115777:	contig of 14579 bp in length
*	115778	134189:	gap of unknown length
*	115778	134189:	contig of 18412 bp in length.

FEATURES  
source  
1. .134189  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-422A19"  
/clone\_11b="RPI mouse BAC library 23"  
BASE COUNT 34661 a 32543 c 32263 g 31734 t 2988 others  
ORIGIN

Query Match 17.8%; Score 196; DB 2; Length 134189;  
Best Local Similarity 100.0%; Pred. No. 2.2e-29;  
Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 aggtcgaaacccagctgcggaaggaagtgccggaagcccgccgagccgagc 102  
|||||  
Db 130269 AGGTCTCAACACGAGCTGGGAGGAGAGTCCGGGACCGCGCAGCCGACG 130210  
|||||  
QY 103 cagggaacggagtgctgtaccagctgcggaacacctgcgccttgcgagcgagc 162  
|||||  
Db 130209 CAGGACACGAGAGTCTGTACCAAGCTGGCGCACACTCTGCCCCCTTTCGGCGCGCGTCCAGC 130150  
|||||  
QY 163 ggcgacacttgaaagagcctccatcatgctgcctacaatacagctaccctgcagtcacgcgc 222  
|||||  
Db 130149 GCGCACCTTGACAGGCGCTCCATCATGCGCTCCACCAATCAGTACTGCGCATGACGCGC 130090  
|||||  
QY 223 ctctcgagcgagcgagtg 238  
|||||  
Db 130089 CTCTGCGCAGCAGGTG 130074  
|||||  
RESULT 15

AF304864	AF304864	3605 bp	mRNA	linear	VF 05-JUN-2001
LOCUS	Oncorhynchus mykiss hypoxia-inducible factor 1 alpha (HIF-1a) mRNA,				
DEFINITION	complete cds.				
ACCESSION	AF304864				
VERSION	AF304864.1	GI:13561505			
KEYWORDS	rainbow trout.				
SOURCE	Oncorhynchus mykiss				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.				
REFERENCE	Soltano,A.J., Rabergh,C.M.I., Gassmann,M., Sistonen,L. and Nikimaa,M.				
AUTHORS	Characterization of a Hypoxia-Inducible Factor (HIF-1 alpha) from Rainbow Trout: Accumulation of Protein Occurs at Normal Venous Oxygen Tension				
TITLE	J. Biol Chem. 276 (23), 19699-19705 (2001)				
JOURNAL	11278461				
PUBMED	2 (bases 1 to 3605)				
REFERENCE	Soltano,A.J., Rabergh,C.M.I., Gassmann,M., Sistonen,L. and Nikimaa,M.				
AUTHORS	Direct Submision				
TITLE	Submitted (12-SEP-2000) Department of Biology, Laboratory of Animal Physiology, University of Turku, Yliopistomakki, Turku FIN-20014, Finland				
JOURNAL					

FEATURES  
source  
1. .3605  
/organism="Oncorhynchus mykiss"  
/db\_xref="taxon:8022"  
1. .3605  
/gene="HIF-1a"  
301. .2601  
/gene="HIF-1a"  
/note="bHLH/PAS A/B transcription factor"  
/codon\_start=1  
/product="hypoxia-inducible factor 1 alpha"  
/protein\_id="AAK30364.1"  
/db\_xref="GI:13561506"  
/translation="MDTGVPPEKRSKVSRRKERSRDAACRGKSEVFEYLADELPLPSVTSNMDKASIMRLAISYLHMNLSITNEEEOERENDSQLNGSYLAIEGFLNVLSEGDWIVLSEVNYKCLGLAIDLTLSVEFTHPCDHELEMLVHRGTSSKRSKEPNTERSFELMKCTLTNRGRTRVNVKSTWVLICSDHVRVHESPAEDIPGGHKPSVPLVLYCDPIRPHSPNTEAPDLTKTFLSRHTIDMKFTYDEKTELKMGDPEDLNRSVVEYTHALDSLDLTKMTHNLFAKGOVSTQYMLAKRGGFVWVETQATVITNNKNSOPQCVGVNVVLSGIIIEBKMLSLQETEDMPVKLEEESSSEPSVPLKKEKSPELDVKLFTTRAVETQPLSLYDRLKEEPEALTLPAAGDTIISLDESSPDSILOKEPILKDWLIPETSDKLALPLSLPPODHVLVNTSVDTPTGSDSSPTGSGSPTEPDSPDRCFPHESDINAEFKIDWETFLFAIRPEKTPPTLOAMEDLDLEMLARYITPMDDFOLRLSPPEPLSCGPAOPLLECSSLSSVRLQEVHSTGSPFNAGSLTASPALNAPSALAAPEPADSPCPASLTKTYPOMDRELISLSLSONQRKRKSLSGAVIGIGLLQDHPGGRKLKVELSESHADAPFNRITLLPLPTDLASRLGLISSBGSPPYLPOLTRYDCEVNAPVGGLQELSELSDADYNA"

BASE COUNT 888 a 951 c 866 g 900 t  
ORIGIN

Query Match 17.5%; Score 192; DB 5; Length 3605;  
Best Local Similarity 59.1%; Pred. No. 7.8e-29;  
Matches 436; Conservative 0; Mismatches 220; Indels 82; Gaps 3;

QY 43 aggtcgaaacccagctgcggaaggaagtgccggaagcccgccgagccgagc 102  
|||||  
Db 334 AGGTGAGCTCGGACCGGAGGAGAGAGTCCGGGATGCGCGGATGACGAGGAGG 393  
|||||  
QY 103 cagggaacggagtgctgtaccagctgcggaacacctgcgccttgcgagcgagtcacgc 162  
|||||  
Db 394 AAGGAATCTGAGGTGTTCTACGAGCTGGGCCAGAGACTGCCCCCTGCGCACACGCTCCAC 453  
|||||  
QY 163 ggcgacacttgaaagagcctccatcatgctgcctacaatacagctaccctgcagtcacgcgc 216  
|||||  
Db 454 TCCACCTTGACAGGCGCTCCATCATGAGACTGGCCATCATGTTACTTTCGACATGAGGAAT 513  
|||||

QY	217	-----caccgcctctgcgcagcagcagtgtaaaaaaggggagagccactggaac	262
Db	514	CTGCTGACACAGACAAATGAGAGGAGACAGAGAGGAGGAAATGTGACTCCACGTGAAT	573
OY	263	gcctgtactacgaaagccctggagaggttctgaagtgtactaccgcgcggaggaacatg	322
Db	574	GGATCTTATCTGMAAGCCTATGAGAGGCTTCTCTATGGTGTCTGTAGAGACGGGACATG	633
OY	323	gcttaacctgtcggaaatgltcaagaaacacacctggagccctagtcagtgagactgttctc	382
Db	634	ATTATCTCTCTGGAAGATGTCAACAAGGCTTGGGCTGTGGCAACG-----	678
OY	383	cctccctgtatatacaaaccccaactcctgltacaaattctctctgagagctattggaaca	442
Db	679	-----ATTGACCTTACAGGACTTA	697
OY	443	gtactcttgatttatcatcatccctgtgtacaaagagagacttaagagccctgaacccca	502
Db	698	GTGTTTGTGAGTACACACACCCCTGTGTGACTGAGGAGCTGTAGAGAACTGTGTACAC	757
OY	503	ggccgaacactgtlcaagaagaagcttggaaagccccaacagacgcaccaatttccctcgaa	562
Db	758	GAACAGGGAGACCTCTAAAAAGTCCAAAGAACCAACACAGAACGAGGCTTCTTGTGGGA	817
OY	563	tgaagaagacgctcaaccaagagagggcgacgctcaactccaagaagcgccactgtgaag	622
Db	818	TGAAATGTACCTCTACCCACMACGGGGGGGCACTGTTCATGTCAATATCGCCACTGGAAAG	877
OY	623	tgcgtcactgtctcaagacatatgagggcctacaagccccctctgaacagaacttcccctccg	682
Db	878	TTTCTCACTGCTCAGACCATGTGGGGGCTCATGTAAATCCCAAGCTGACGACAGATCCCGGTG	937
OY	683	ggagccctgcgtccgagagcctcccttgaaatgactgtgtgtactatcttgyaaagcatccccc	742
Db	938	GG-----CAACAAGAGGCGCGCTCCCTCCCACTCGGAGGTGTGTGTGTGACCCCATCCCC	991
OY	743	agctcccttccaagatg	760
Db	992	ACCCCTCCACATCGAGG	1009

Search completed: September 8, 2002, 07:29:17  
Job time: 6900 sec

•  
•  
•  
•

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2002, 05:58:42 ; Search time 49.21 Seconds  
(without alignments)  
692.942 Million cell updates/sec

Title: US-09-896-791B-3

Perfect score: 1636

Sequence: 1 MALIGNORVSNTELRKEKSR.....TESSLPSTWVLMALNRKNCPS 307

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	855.5	52.3	662	20	AAV06295
2	831	50.8	632	22	AAB93326
3	573	35.0	648	22	AAB94934
4	518.5	31.7	199	22	AAB93710
5	495	30.3	245	21	AAV94627
6	495	30.3	330	21	AAV94628
7	495	30.3	642	21	AAV84168
8	495	30.3	652	21	AAV94629
9	495	30.3	669	21	AAV84167
10	495	30.3	697	21	AAV84166
11	495	30.3	701	21	AAV84173

12	495	30.3	710	21	AAV84172
13	495	30.3	724	21	AAV84171
14	495	30.3	749	21	AAV84170
15	495	30.3	789	21	AAV84169
16	495	30.3	813	21	AAV94636
17	495	30.3	826	18	AAV06557
18	495	30.3	826	19	AAV80418
19	495	30.3	826	20	AAV06289
20	495	30.3	826	20	AAV94640
21	495	30.3	826	21	AAV69407
22	495	30.3	826	21	AAV69407
23	477.5	29.2	875	19	AAV06298
24	467	28.5	870	20	AAV06290
25	466	28.5	485	21	AAV79160
26	466	28.5	870	19	AAV37097
27	466	28.5	870	21	AAV79161
28	444	27.1	373	18	AAV06559
29	444	27.1	805	18	AAV06558
30	426	26.0	237	22	AAV95649
31	345.5	21.1	613	21	AAV94630
32	323	19.7	373	22	ABG20871
33	323	19.7	756	21	AAV94635
34	294	18.0	590	19	AAV68091
35	293.5	17.9	594	19	AAV68092
36	281	17.2	1507	22	ABV58562
37	278.5	17.0	691	20	AAV06294
38	269	16.4	776	22	ABG28236
39	265	16.2	634	22	ABV71624
40	264	16.2	657	18	AAV30758
41	264	16.1	828	22	ABG20868
42	251	15.3	958	22	ABV62764
43	250.5	15.3	924	21	AAV52991
44	222.5	13.6	99	21	AAV54411
45	214	13.1	74	21	AAV94626

#### ALIGNMENTS

RESULT 1	
AAV06295	
AAV06295 standard; Protein: 662 AA.	
XX	
AC	AAV06295;
XX	
DT	23-AUG-1999 (first entry)
XX	
DE	Mouse transcription regulator MOP7.
XX	
KW	MOP7; member of the PAS superfamily; bHLH-PAS; mouse;
KW	transcription regulator; hypoxia inducible factor 3 alpha.
XX	
OS	Mus musculus.
XX	
PN	MO9928464-A2.
XX	
PD	10-JUN-1999.
XX	
PF	27-NOV-1998; 98WO-US25314.
XX	
PR	28-NOV-1997; 97US-0066863.
XX	
PA	(WISC ) WISCONSIN ALUMNI RES FOUND.
XX	
PI	Bradfield CA, Gu YZ, Hogenesch JB;
XX	
DR	WPI: 1999-371120/31.
XX	
XX	N-PSDB: AAX58986.
PT	Developmental signal transduction associated proteins
XX	
PS	Claim 6; Page 101; 106pp; English.
XX	

A variant of human  
A variant of human  
A variant of human  
A variant of human  
HIF-1alpha variant  
Human hypoxia indu  
Amino acid sequenc  
Human transcriptio  
Human hypoxia-indu  
A wild type human  
Human lung tumor  
Murine endothelial  
Human transcriptio  
Human endothelial  
Human endothelial  
HIF-1 alpha delta-  
Hypoxia inducible  
Human protein sequ  
HIF-1alpha variant  
Novel human diagno  
HIF-1alpha variant  
Human neuronal PAS  
Mouse neuronal PAS  
Drosophila melanog  
Human transcriptio  
Novel human diagno  
Drosophila melanog  
Murine SIM (single  
Novel human diagno  
Drosophila melanog  
Drosophila melanog  
Human pancreatic c  
Hypoxia-inducible

CC The present sequence represents mouse MOP7, a novel member of the  
CC PAS superfamily, where PAS stands for PER/ARNT/SIM domains. MOP7  
CC cDNA (see AAX58986) was identified in a search of murine ESTs designed  
CC to identify bHLH-PAS proteins, and by RACE amplification of lung  
CC cDNA. MOP7 was characterised as hypoxia inducible factor 3 alpha  
CC (HIF 3 alpha). Its expression profile is distinct from that of  
CC HIF 1 alpha (see AY06289), HIF 2 alpha (see AAU06290), MOP3 (see  
CC AY06291), Ah receptor and Ah receptor nuclear translocator (ARNT),  
CC suggesting a different functional role. MOP7 probably regulates  
CC the same genes as HIF 1 alpha and 2 alpha, as evidenced by its  
CC dimerisation with the same partners (ARNT, MOP3) and recognition  
CC of the same core response element. MOP7 may have a functional  
CC role associated with response to low oxygen in the tissues in  
CC which it is expressed. The invention provides novel MOPs 2-9.  
CC nucleic acids (see AAX58981-88) and proteins (see AY06289-97).  
CC These are useful in a variety of research, diagnostic and  
CC therapeutic applications. Several of the MOPs are alpha-class  
CC hypoxia-inducible factors. Others are involved in circadian signal  
CC transduction.

XX Sequence 662 AA;

Query Match 52.3%; Score 855.5; DB 20; Length 662;  
Best Local Similarity 68.4%; Pred. No. 7.6e-80; Indels 55; Gaps 4;  
Matches 182; Conservative 5; Mismatches 24;

YY 9 RSMTELKRSRDAARSRQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHR 68  
DB 7 rsntelrkrsrdaarrrrsqetevlyqlahltlpfargvsahldkasimrltisylylmhr 66  
YY 69 LCAAGKRGATGRLPEPGGFRHGRHGRHGLPVKCCQDAPGQSDVC----- 120  
DB 67 lcaagewn-----qvekgyepdladcykalegf 94  
YY 121 -----SSSLIHNPTGNTFS-LELIGSTPDTFHPCQDELOALPRMULSKKL 170  
DB 95 vmyltaegdmaylseyvskhlglsqlelghnsifdihpcdeeqdaltpripnlsskkl 154  
YY 171 EAAPERHFSLRMKSSTLSRGRTLNKAATWKLHSCGHRAYKPPAQTSPAGSPRSEPL 230  
DB 155 eaaperhfslrmskstlsrgrtlukaatwkvlnhscghmrarxkppaqtspagspsepl 214  
YY 231 QCLVLICEAIP-----QLPFHDGATL 251  
DB 215 qclvlliceaiphslepplgrgafll 240

RESULT 2

AAB93326 standard; Protein: 632 AA.

XX AAB93326;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12422.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 12422; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-PT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 632 AA;

Query Match 50.8%; Score 831; DB 22; Length 632;  
Best Local Similarity 68.2%; Pred. No. 2.5e-77;  
Matches 178; Conservative 11; Mismatches 30; Indels 42; Gaps 5;

YY 1 MALGLQVRNSNTELRKESRDARSRSOETEVLYQLAHTLPFARGVSAHLDKASIMRLT 60  
DB 1 malglqrarsttelrkrksrdaarrrrsqetevlyqlahltlpfargvsahldkasimrlt 60

YY 61 ISYLRMHRLLCAAGKRGATGRLPEPGGFRHGRHGRHGLPVKCCQ----- 109  
DB 61 isylrmhrllcaag-----ewngvgaggepldacykalegfymvl 100

YY 110 QAPGQSDVCSSSLIHNPTGNTFS-LELIGSTPDTFHPCQDELOALPRMULSKKL 168  
DB 101 taegdmay-----lseyvskhlglsqlelghnsifdihpcdeeqdaltpqgltssr 154

YY 169 KLEAPERHFSLRMKSSTLSRGRTLNKAATWKLHSCGHRAYKPPAQTSPAGSPRSEPL 228  
DB 155 kveaperhfslrmskstlsrgrtlukaatwkvlnhscghmrarxkppaqtspagspsepl 214

YY 229 PLQCLVLICEAIPQPLPHDGA 249  
DB 215 plqclvlliceaip-----hpgs 231

RESULT 3

AAB94934 standard; Protein: 648 AA.

XX AAB94934;

XX 26-JUN-2001 (first entry)

XX 26-JUN-2001 (first entry)

XX 26-JUN-2001 (first entry)

XX 26-JUN-2001 (first entry)

XX 26-JUN-2001 (first entry)



```

Query Match          31.7%; Score 518.5; DB 22; Length 199;
Best Local Similarity 68.0%; Pred. No. 1.5e-45;
Matches 104; Conservative 9; Mismatches 13; Indels 27; Gaps 2

OY 120 CSSSLIHNPPTPTNF-----SLEILGHSIFPIHPDCDEIQ 156
Db 14 ctswtltropspasapptwtrlpiscaspsatcctaapqlelghsifdfihpcdeeq 73

OY 157 DALPRPNLSKKRLKAPERHFSLRKSTLTLSRGRTLNKAAWTWYKHSCHRAVAPPA 216
Db 74 daltppqtlstrkveaptercftsmkstltstrgrltnlkaatkwlnvncsglmraykppa 133

OY 217 QTSFAGSPRSEPPLOCVLVLCALFQLPERHQA 249
Db 134 qlspagsposepplqlcylvicaalp----hps 162

RESULT 5
AAY94627
ID AAY94627 standard; protein; 245 AA.
AC AAY94627;
XX
XX
XX 15-AUG-2000 (first entry)
DE HIF-1alpha variant protein sequence HIF-1alpha/1-245.
XX
XX Hypoxia-inducible factor 1alpha; HIF-1alpha; PAS-B; N-TAD; C-TAD;
KW regulation; angiogenesis; erythropoiesis; glycolysis; human.
XX
XX Homo sapiens.
OS
XX WO200029437-A1.
PN
XX 25-MAY-2000.
PD
XX 11-NOV-1999; 99WO-SE02053.
PE
XX 13-NOV-1998; 98SE-0003891.
PR
XX (PHAA ) PHARMACIA & UPJOHN AB.
PA
XX Berkenstam A, Poellinger L;
XX
XX WPI; 2000-399715/34.
DR
XX
XX
XX Human hypoxia-inducible factor alpha variants for identifying compounds
PT that modulate its functional domain and regulate genes involved in
PT angiogenesis, erythropoiesis -
PS
XX
XX Claim 13; Page 68-69; 87pp; English.
XX
XX This sequence represents a fragment of the hypoxia-inducible factor
XX (HIF)-1alpha amino acid sequence. The mechanism of action of HIF-1alpha
XX is a multi-step process which includes hypoxia-dependent nuclear import
XX and activation of the transactivation domain. The HIF-1alpha consists of
XX a number of functional domains including a PAS-B (Per, Arnt, Sim) domain
XX located in human HIF-1alpha between amino acids 173 and 390, a C-terminal
XX nuclear localization sequence located at amino acids 718-584, a
XX transactivator domain (N-TAD) located between amino acids 531 and 584,
XX and a second transactivator domain (C-TAD) located between 813 and 826
XX The invention relates to isolated variants of HIF-1alpha, such as that
XX represented by the present sequence. The variants are useful for
XX identifying compounds capable of modulating the function of a functional
XX domain of human HIF-1alpha. The method comprises contacting a candidate
XX compound with a cell expressing a HIF-1alpha variant conjugated to a
XX molecular probe. The localization of the probe can be detected in the
XX cell. The Aqueorea victoria green fluorescent protein can be used as the
XX molecular probe. The compounds are useful for the regulation of
XX HIF-1alpha target genes, such as those involved in the regulation of
XX angiogenesis, erythropoiesis and glycolysis.
XX
XX Sequence 245 AA:

```

Query Match	30.3%	Score 495;	DB 21;	Length 245;
Best Local Similarity	46.2%	Pred. No.5.6e-43;		
Matches 114;	Conservative 26;	Mismatches 59;	Indels 48;	Gaps 6
QY	11	NTELRKESGRDAARSRSHQSEFEVLQIATLPFARGVSAHLKASIMRLTISYLRHRLC	70	
DB	14	ASGRKESKRDAAARSRRSHQSEFEVLQIATLPFARGVSAHLKASIMRLTISYLRHRLC	73	
QY	71	AAAGKRGARATGRLLPEGPGGFRHGHRRGRHGLPVGKCOOA-----PEPOSVDLC	120	
DB	74	dag-----diededamkagmncfykaldgfvnwld 105		
QY	121	SSSLIH-----NPTREG-TNFSLELIGHSTFDFPHPDQDELDADLPRLNLSKKKLEAPR	174	
DB	106	dgdmdyidsavhkymlqgf-elqhsyvdftchpdcbeemcmemlthrglykkykqent	163	
QY	175	EGHFSLRMRKSTLTLSNRGRTLNLRKAATKVKVLCGSHMRVAKRPPAQTSPAGSPRSEPPQLQCLV	234	
DB	164	qrsfflmkccltsrgtrtmiksatkvlhcgghlvdyutnsnqpqcg--ykkkppmtclv	221	
QY	235	LICEALP 241		
DB	222	licealp 228		
RESULT	6			
AA94628				
ID	AA94628	standard; protein; 330 AA.		
XX	AA94628;			
XX	15-AUG-2000	(first entry)		
DE	HIF-1alpha variant protein sequence HIF-1alpha/1-330.			
XX	Hypoxia-inducible factor 1alpha; HIF-1alpha; PAS-B; N-TAD; C-TAD;			
KW	regulation; angiogenesis; erythropoiesis; glycolysis; human.			
XX	Homo sapiens.			
PN	WO200029437-A1.			
PD	25-MAY-2000.			
XX	11-NOV-1999;	99WO-SE02053.		
PF	13-NOV-1998;	98SE-0003891.		
XX	(PAAA ) PHARMACIA & UPJOHN AB.			
PA	Berkeinstam A, Poellinger L;			
PI	WPI: 2000-399715/34.			
DR	Human hypoxia-inducible factor alpha variants for identifying compounds			
XX	that modulate its functional domain and regulate genes involved in			
PT	angiogenesis -			
PS	Claim 15; Page 69-70; 87Pp; English.			
XX	This sequence represents a fragment of the hypoxia-inducible factor			
CC	(HIF)-1alpha amino acid sequence. The mechanism of action of HIF-1alpha			
CC	is a multi-step process which includes hypoxia-dependent nuclear import			
CC	and activation of the transactivation domain. The HIF-1alpha consists of			
CC	a number of functional domains including a PAS-B (Per, Arnt, Sim) domain			
CC	located in human HIF-1alpha between amino acids 173 and 390, a C-terminal			
CC	nuclear localization sequence located at amino acids 718-584, a			
CC	transactivator domain (N-TAD) located between amino acids 531 and 584,			
CC	and a second transactivator domain (C-TAD) located between 813 and 826.			
CC	The invention relates to isolated variants of HIF-1alpha, such as that			
CC	represented by the present sequence. The variants are useful for			

PT Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for  
PT treating hypoxia or ischemia-related tissue damage -  
XX  
PS Claim 1, Page -; 96pp; English.

XX (PMAA ) PHARMACIA & UPJOHN AB.  
PA  
XX  
PI Berkenstam A, Poellinger L;





DT		03-JUL-2000	(first entry)
DE		A variant of human hypoxia inducible factor-1 alpha protein.	
KW		Human; hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;	
KW		hypoxia inducible gene; hypoxia inducible factor; hypoxia;	
KW		ischemia related damage; angiogenesis; coronary artery disease;	
XK		ischemic tissue damage.	
OS		Synthetic.	
OS		Homo sapiens.	
FT	Key	Location/Qualifiers	
FT	Misc-difference 422	/note= "this residue is optionally not Ser, and is preferably Gly"	
FT	Misc-difference 423	/note= "this residue is optionally not Thr, and is preferably Ala"	
XX		WO200010578-A1.	
PX		02-MAR-2000.	
PX		25-AUG-1999; 99WO-US19416.	
PX		25-AUG-1998; 98US-0148547.	
PA		(UOJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.	
PL		Semenza GL;	
DR		WPI; 2000-246493/21.	
PT		Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for treating hypoxia or ischemia-related tissue damage -	
PS		Claim 1; Page -: 96pp; English.	
CC		The present sequence represents a variant of hypoxia-inducible factor (HIF)-1 alpha, comprising amino acids 1-391 and 521-826 of the wild type protein (see AY65407). The HIF-1alpha variants are stable under hypoxic and non-hypoxic conditions. The variants comprises amino acid residues 1-391 and 521-826, 549-826, 576-826, 429-826, 469-826, 494-826, 508-826, 512-826 or 517-826 of the wild type human HIF-1alpha polypeptide, in which residues 551 and 552 are not serine and threonine, respectively. The HIF-1alpha variant polynucleotide sequences are useful for increasing expression of a hypoxia inducible gene in a cell. They is also useful for providing constitutive expression of a hypoxia inducible factor in a cell, and for reducing or preventing hypoxia or ischemia related damage. The variant HIF-1alpha polypeptides are useful for providing prophylactic therapy for inducing the level of angiogenesis in tissues of patients at risk of coronary artery disease or ischemic tissue damage.	
CC		note: this sequence does not appear in the specification; it was created using information provided.	
XX		Sequence 697 AA:	
OY		Query Match 30.3%; Score 495; DB 21; Length 697; Best Local Similarity 46.2%; Pred. No. 2.5e-42; Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;	
DB		11 NTELRKESROAARSRSQETEVLYQLAHNTLPFARGVSAHLDKASIMRLTISYLMMNRIC 70 ::  14 ssetrkeksidaarstrkesevfyelabqjrlphnvsshldkasvmtltisyltlvrkll 73	
OY		71 AAGGKGATGATGLPEGGGFRNTHNGRNIGLPRGKCGQA-----PGPSYDLC 120  DB 74 day-----dlidedmkaqmncfylvkalidgfmwvltd 105	
OY		121 SSSLIH-----NPRTG-TNFSLLEIGHSIFFPHPCDOEEIOLADLTTPRNISSKKYLEAP 174	

Dd		dgdmiysadnvnkymgltgf--eltgnsvdfdtcpchdeemremcltnrglrvkkqegnt	163
Oy	175	ERHPSLRMKSTLTSTRGRGTLNLKATWVFLHCSGHMRAVKPPAQTSPAGSRSEPPLQCLV	234
Dd	164	grsflltmkccltsrgrtmtmkstkwvlvtcgtghlvhydctnsqpqcg--ykkprptmcltv	221
Oy	235	LICEAIP	241
Dd	222	liceip	228
<hr/>			
RESULT 11			
AAY84173	ID	AAY84173 standard; Protein; 701 AA.	
AAV84173:	AC		
03-JUL-2000	(first entry)		
A variant of human hypoxia inducible factor-1 alpha protein.			
Human: hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;			
KM		hypoxia inducible gene; hypoxia inducible factor; hypoxia;	
KW		ischemia related damage; angiogenesis; coronary artery disease;	
KW		ischemic tissue damage.	
XX		Synthetic.	
OS		Homo sapiens.	
XX			
FH		Location/Qualifiers	
FT	Misc-difference	426	
FT	/note=	"this residue is not Ser, and is preferably Gly"	
FT	Misc-difference	427	
FT	/note=	"this residue is not Thr, and is preferably Ala"	
PN	WC2000I0578-A1.		
PD	02-MAR-2000.		
XA	25-AUG-1999;	99WO-US19416.	
XF	25-AUG-1998;	98US-0148547.	
PA	(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.		
P1	Semenza GL;		
DR	WPI; 2000-246493/21.		
PT	Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for treating hypoxia or ischemia-related tissue damage -		
PS	Claim 1; Page -: 96pp; English.		
CC	The present sequence represents a variant of hypoxia-inducible factor (HIF)-1 alpha, comprising amino acids 1-391 and 517-826 of the wild type protein (see AAY69407). The HIF-1alpha variants are stable under hypoxic and non-hypoxic conditions. The variants comprises amino acid residues 1-391 and 521-826, 549-826, 576-826, 469-826, 494-826, 508-826, 512-826 or 517-826 of the wild type human HIF-1alpha polypeptide, in which residues 551 and 552 are not serine and threonine, respectively. The HIF-1alpha variant polynucleotide sequences are useful for increasing expression of a hypoxia inducible gene in a cell. They are also useful for providing constitutive expression of a hypoxia inducible factor in a cell, and for reducing or preventing hypoxia or ischemia related damage. The variant HIF-1alpha polypeptides are useful for providing prophylactic therapy for inducing the level of angiogenesis in tissues of patients at risk of coronary artery disease or ischemic tissue damage.		
CC	note: This sequence does not appear in the specification: It was created by computer from a patent application filed by Dr. Michael J. Semenza et al., dated 1997-05-16, assigned to Genzyme Corporation, Boston, MA, USA.		

CC using information provided.  
XX  
XX Sequence 701 AA;

Query Match 30.3%; Score 495; DB 21; Length 701;  
Best Local Similarity 46.2%; Pred. No. 2.5e-42;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;

QY 11 NTELRKESRDAAARRSRROEYVYOLAHLPFARGVSAHLDKASIMRLTSTYLRMRIC 70  
DB 14 sserikksrdaarsrskesevylahqlplphvsshldkasmrltistylrvkll 73  
QY 71 AAGGKRGRATGRLLPEGGPGRHGTNRGRHGLPVGKCOQA-----PGQSYDLC 120  
DB 74 dag-----dlidddmkagmcfylkaldfvmvltld 105  
QY 121 SSSLIIH-----NPPPG-TNSLELIGHSTFDTHPCQDELDALPRPMLSKKLEAPT 174  
DB 106 dgdmllyisdhvnkymgtlqf--eltghsvdfthpcdheemremthrnglvkkgkeqnt 163  
QY 175 ERHPSLRKSTLTSRGRTLNLKATWVLHCSGHRAYKPPAQTSPAGSPRSEPPLOCIV 234  
DB 164 grsfflmkctllsrgtltmiksawkvlnctghihvydtnsgpqcg--ykkpmtciv 221  
QY 235 LICEAIP 241  
DB 222 lliceip 228

RESULT 12  
AAy84172  
XX ID AAY84172 standard; Protein; 710 AA.  
AC AAY84172;  
XX  
XX 03-JUL-2000 (first entry)  
DE A variant of human hypoxia inducible factor-1 alpha protein.  
XX  
XX Human: hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;  
KW hypoxia inducible gene; hypoxia inducible factor; hypoxia;  
KW ischemia related damage; angiogenesis; coronary artery disease;  
KW ischemic tissue damage.  
XX  
XX Synthetic.  
OS Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 435 /note- "this residue is not Ser, and is  
FT /note- "this residue is not Ser, and is  
FT Misc-difference 436 preferably Gly"  
FT /note- "this residue is not Thr, and is  
FT /note- "this residue is not Thr, and is  
FT Preferably Ala"  
XX  
XX MO200010578-A1.  
XX  
XX PN  
XX DE  
XX PD 02-MAR-2000.  
XX PF 25-ANG-1999; 99MO-US19416.  
XX PR 25-AUG-1998; 98US-0148547.  
XX  
XX PA (UYJO ) UNITV JOHNS HOPKINS SCHOOL MEDICINE.  
XX  
XX PI Semenza GL;  
XX  
XX DR WPI; 2000-246493/21.  
XX  
XX PT Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for  
PT treating hypoxia or ischemia-related tissue damage -  
XX  
XX FT

PS Claim 1; Page -: 96pp; English.  
XX  
XX The present sequence represents a variant of hypoxia-inducible factor  
CC (HIF)-1 alpha, comprising amino acids 1-391 and 508-826 of the wild  
CC type protein (see AAY9407). The HIF-1alpha variants are stable under  
CC hypoxic and non-hypoxic conditions. The variants comprises amino acid  
CC residues 1-391 and 521-826, 549-826, 576-826, 429-826, 469-826,  
CC 508-826, 512-826 or 517-826 of the wild type human HIF-1alpha  
CC polypeptide, in which residues 551 and 552 are not serine and threonine,  
CC respectively. The HIF-1alpha variant polynucleotide sequences are useful  
CC for increasing expression of a hypoxia inducible gene in a cell. They  
CC are also useful for providing constitutive expression of a hypoxia  
CC inducible factor in a cell, and for reducing or preventing hypoxia or  
CC ischemia related damage. The variant HIF-1alpha polypeptides are useful  
CC for providing prophylactic therapy for inducing the level of  
CC angiogenesis in tissues of patients at risk of coronary artery disease  
CC or ischemic tissue damage.  
CC note: this sequence does not appear in the specification; it was created  
CC using information provided.  
XX  
XX SQ Sequence 710 AA;

Query Match 30.3%; Score 495; DB 21; Length 710;  
Best Local Similarity 46.2%; Pred. No. 2.6e-42;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;

QY 11 NTELRKESRDAAARRSRROEYVYOLAHLPFARGVSAHLDKASIMRLTSTYLRMRIC 70  
DB 14 sserikksrdaarsrskesevylahqlplphvsshldkasmrltistylrvkll 73  
QY 71 AAGGKRGRATGRLLPEGGPGRHGTNRGRHGLPVGKCOQA-----PGQSYDLC 120  
DB 74 dag-----dlidddmkagmcfylkaldfvmvltld 105  
QY 121 SSSLIIH-----NPPPG-TNSLELIGHSTFDTHPCQDELDALPRPMLSKKLEAPT 174  
DB 106 dgdmllyisdhvnkymgtlqf--eltghsvdfthpcdheemremthrnglvkkgkeqnt 163  
QY 175 ERHPSLRKSTLTSRGRTLNLKATWVLHCSGHRAYKPPAQTSPAGSPRSEPPLOCIV 234  
DB 164 grsfflmkctllsrgtltmiksawkvlnctghihvydtnsgpqcg--ykkpmtciv 221  
QY 235 LICEAIP 241  
DB 222 lliceip 228

RESULT 13  
AAy84171  
XX ID AAY84171 standard; Protein; 724 AA.  
AC AAY84171;  
XX  
XX 03-JUL-2000 (first entry)  
DE A variant of human hypoxia inducible factor-1 alpha protein.  
XX  
XX Human: hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;  
KW hypoxia inducible gene; hypoxia inducible factor; hypoxia;  
KW ischemia related damage; angiogenesis; coronary artery disease;  
KW ischemic tissue damage.  
XX  
XX Synthetic.  
OS Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 449 /note- "this residue is not Ser, and is  
FT /note- "this residue is not Ser, and is  
FT Misc-difference 450 preferably Gly"  
FT /note- "this residue is not Thr, and is  
FT /note- "this residue is not Thr, and is  
FT Preferably Ala"  
XX  
XX FT

```
XX
PN WO200010578-A1.
XX
XX 02-MAR-2000.
XX
XX 25-AUG-1999; 99WO-US19416.
XX
XX 25-AUG-1998; 98US-0148547.
XX
XX (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
XX Semenza GL;
XX
XX WPI; 2000-246493/21.
XX
XX Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for
XX treating hypoxia or ischemia-related tissue damage .
XX
XX Claim 1; Page -: 96pp; English.
XX
XX The present sequence represents a variant of hypoxia-inducible factor
XX (HIF)-1 alpha, comprising amino acids 1-391 and 494-826 of the wild
XX type protein (see AAY69407). The HIF-1alpha variants are stable under
XX hypoxic and non-hypoxic conditions. The variants comprises amino acid
XX residues 1-391 and 521-826, 549-826, 429-826, 469-826, 494-826,
XX 508-826, 512-826 or 517-826 of the wild type human HIF-1alpha
XX polypeptide, in which residues 551 and 552 are not serine and threonine,
XX respectively. The HIF-1alpha variant polynucleotide sequences are useful
XX for increasing expression of a hypoxia inducible gene in a cell. They
XX is also useful for providing constitutive expression of a hypoxia
XX inducible factor in a cell, and for reducing or preventing hypoxia or
XX ischemia related damage. The variant HIF-1alpha polypeptides are useful
XX for providing prophylactic therapy for inducing the level of
XX angiogenesis in tissues of patients at risk of coronary artery disease
XX or ischemic tissue damage.
XX note: this sequence does not appear in the specification; it was created
XX using information provided.
XX
XX Sequence 724 AA;
XX
XX Query Match 30.3%; Score 495; DB 21; Length 724;
XX Best Local Similarity 46.2%; Pred. No. 2.7e-42;
XX Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
XX
XX 11 NTELKESRDAARSRROETVLYQLAHTLPFANGVSAHLDKASIMRLTISYLRMRIC 70
XX :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :|
XX 14 sserfkksrdaarsrskesevfylahqjrlphnvshldkasvmtltsylivrkll 73
XX
XX 71 AAGGKRGATGRLRPEGGGRNHTNRGRHGLRPGKCOQA-----PGQSDYDC 120
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 74 dag-----dlldedmkqgmncfykaldgfymvltld 105
XX
XX 121 SSSLIH-----NPTPG-TNFSLELIGHSTFDFINPCDELODALTPRPNLSKKKLEAPT 174
XX :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 106 dgdmyisdnykylnglqf--elcgshvfditphcdheememltlhnglvkxkqent 163
XX
XX 175 EHNFLBMRKSTLTSRGRTLNKKAATWVKVLCGSHMRKATPRPQTSBPASSPSEPPLOCTY 234
XX :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 164 qrsflrmkcltstgrtmniksalwkvlyhctghlvhdtnsnpgcg--ytkppmtctiv 221
XX
XX 235 LICEAIP 241
XX |||| ||
XX 222 lliccpip 228
XX
XX RESULT 14
XX AAY84170
XX ID AAY84170 standard; Protein: 749 AA.
XX
XX AAY84170;
XX
XX 03-JUL-2000 (first entry)
```

```
XX
DE A variant of human hypoxia inducible factor-1 alpha protein.
XX
XX Human; hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;
XX hypoxia inducible gene; hypoxia inducible factor; hypoxia;
XX ischemia related damage; angiogenesis; coronary artery disease;
XX ischemic tissue damage.
XX
XX Synthetic.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 474
XX FT /note= "this residue is not Ser, and is
XX FT /note= "preferably Gly"
XX FT Misc-difference 475
XX FT /note= "this residue is not Thr, and is
XX FT /note= "preferably Ala"
XX
XX WO200010578-A1.
XX
XX 02-MAR-2000.
XX
XX 25-AUG-1999; 99WO-US19416.
XX
XX 25-AUG-1998; 98US-0148547.
XX
XX (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX Semenza GL;
XX
XX WPI; 2000-246493/21.
XX
XX Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for
XX treating hypoxia or ischemia-related tissue damage .
XX
XX Claim 1; Page -: 96pp; English.
XX
XX The present sequence represents a variant of hypoxia-inducible factor
XX (HIF)-1 alpha, comprising amino acids 1-391 and 469-826 of the wild
XX type protein (see AAY69407). The HIF-1alpha variants are stable under
XX hypoxic and non-hypoxic conditions. The variants comprises amino acid
XX residues 1-391 and 521-826, 549-826, 429-826, 469-826, 494-826,
XX 508-826, 512-826 or 517-826 of the wild type human HIF-1alpha
XX polypeptide, in which residues 551 and 552 are not serine and threonine,
XX respectively. The HIF-1alpha variant polynucleotide sequences are useful
XX for increasing expression of a hypoxia inducible gene in a cell. They
XX is also useful for providing constitutive expression of a hypoxia
XX inducible factor in a cell, and for reducing or preventing hypoxia or
XX ischemia related damage. The variant HIF-1alpha polypeptides are useful
XX for providing prophylactic therapy for inducing the level of
XX angiogenesis in tissues of patients at risk of coronary artery disease
XX or ischemic tissue damage.
XX note: this sequence does not appear in the specification; it was created
XX using information provided.
XX
XX Sequence 749 AA;
XX
XX Query Match 30.3%; Score 495; DB 21; Length 749;
XX Best Local Similarity 46.2%; Pred. No. 2.8e-42;
XX Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
XX
XX 11 NTELKESRDAARSRROETVLYQLAHTLPFANGVSAHLDKASIMRLTISYLRMRIC 70
XX :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 14 sserfkksrdaarsrskesevfylahqjrlphnvshldkasvmtltsylivrkll 73
XX
XX 71 AAGGKRGATGRLRPEGGGRNHTNRGRHGLRPGKCOQA-----PGQSDYDC 120
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 74 dag-----dlldedmkqgmncfykaldgfymvltld 105
XX
XX 121 SSSLIH-----NPTPG-TNFSLELIGHSTFDFINPCDELODALTPRPNLSKKKLEAPT 174
XX :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: September 8, 2002, 06:45:32 : Search time 19.62 Seconds  
(without alignments)  
382.195 Million cell updates/sec

Title: US-09-896-791B-3  
Perfect score: 1636  
Sequence: 1 MALGIGRVNSNTELRKEKSR.....TESSLSPTWVLMALRNKCPG 307

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCYUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	495	30.3	826	1 US-08-785-241-6	Sequence 6, Appl1
2	495	30.3	826	2 US-08-480-473B-2	Sequence 2, Appl1
3	495	30.3	826	3 US-08-915-213-2	Sequence 2, Appl1
4	495	30.3	826	3 US-09-148-547-2	Sequence 2, Appl1
5	495	30.3	826	4 US-09-235-217-2	Sequence 2, Appl1
6	495	30.3	826	5 PCT-US96-10251-2	Sequence 2, Appl1
7	489	29.9	810	1 US-08-785-241-7	Sequence 7, Appl1
8	477.5	29.2	875	1 US-08-785-241-5	Sequence 5, Appl1
9	466	28.5	870	1 US-08-785-241-4	Sequence 4, Appl1
10	444	27.1	373	2 US-08-480-473B-3	Sequence 3, Appl1
11	444	27.1	373	3 US-08-915-213-3	Sequence 3, Appl1
12	444	27.1	373	4 US-09-235-217-3	Sequence 3, Appl1
13	444	27.1	373	5 PCT-US96-10251-3	Sequence 3, Appl1
14	444	27.1	805	2 US-08-480-473B-4	Sequence 4, Appl1
15	444	27.1	805	3 US-08-915-213-4	Sequence 4, Appl1
16	444	27.1	805	4 US-09-235-217-4	Sequence 4, Appl1
17	444	27.1	805	5 PCT-US96-10251-4	Sequence 4, Appl1
18	294	18.0	590	2 US-08-785-310A-5	Sequence 5, Appl1
19	293.5	17.9	594	2 US-08-480-473B-36	Sequence 6, Appl1
20	203	12.4	54	2 US-08-915-213-36	Sequence 36, Appl1
21	203	12.4	54	3 US-09-235-217-36	Sequence 36, Appl1
22	203	12.4	54	4 US-09-150-460B-17	Sequence 17, Appl1
23	197	12.0	50	4 US-09-150-460B-18	Sequence 18, Appl1
24	188	11.5	50	4 US-09-150-460B-15	Sequence 9, Appl1
25	142.5	8.3	789	4 US-08-971-188-9	Sequence 38, Appl1
26	136.5	8.3	789	4 US-08-971-188-9	Sequence 38, Appl1
27	131	8.0	53	2 US-08-480-473B-38	Sequence 38, Appl1

28	131	8.0	53	3 US-08-915-213-38	Sequence 38, Appl1
29	131	8.0	53	4 US-09-235-217-38	Sequence 38, Appl1
30	125.5	7.7	626	4 US-08-971-188-10	Sequence 10, Appl1
31	124.5	7.6	51	4 US-09-150-460B-16	Sequence 16, Appl1
32	120.5	7.4	631	4 US-08-971-188-8	Sequence 8, Appl1
33	107	6.5	102	2 US-08-480-473B-47	Sequence 47, Appl1
34	107	6.5	102	3 US-08-915-213-47	Sequence 47, Appl1
35	107	6.5	102	4 US-09-235-217-47	Sequence 47, Appl1
36	106	6.5	808	4 US-08-971-188-12	Sequence 12, Appl1
37	106	6.5	848	1 US-08-045-806-4	Sequence 4, Appl1
38	106	6.5	848	1 US-08-366-051B-4	Sequence 4, Appl1
39	103	6.3	716	4 US-08-971-188-5	Sequence 5, Appl1
40	102	6.2	716	4 US-08-971-188-11	Sequence 11, Appl1
41	102	6.2	737	4 US-08-971-188-4	Sequence 4, Appl1
42	99	6.1	501	2 US-08-660-963-13	Sequence 13, Appl1
43	91.5	5.6	846	3 US-08-885-291-55	Sequence 55, Appl1
44	91.5	5.6	846	3 US-09-107-847-2	Sequence 2, Appl1
45	91.5	5.6	846	4 US-09-496-672-55	Sequence 55, Appl1

ALIGNMENTS

```
RESULT 1
US-08-785-241-6
: Sequence 6, Application US/08785241
: Patent No. 5695963
:
: GENERAL INFORMATION:
:
: APPLICANT: McKnight, Steven L.
: APPLICANT: Russell, David W.
: APPLICANT: Tian, Hui
:
: TITLE OF INVENTION: Endothelial PAS Domain Protein
:
: NUMBER OF SEQUENCES: 7
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 268 BUSH STREET, SUITE 3200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/785,241
: FILING DATE: 17-JAN-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UTSD:1229
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 343-4341
: TELEFAX: (415) 343-4342
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 826 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-785-241-6
```

Query Match 30.3% Score 495; DB 1; Length 826;  
Best Local Similarity 46.2% Pred. No. 1.1e-45;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;  
OY 11 NTELRKEKSRPAASRRSOETFEVLYOLAHLPFGAGVSAHLDDKASIMPLTISYLEMRRLC 70  
DB 14 SSERKREKSRPAASRRKESEVFEIHLAQLPLPHNVSHLDDKASVMTLTISYLEMRRL 73

```
OY 71 AAGKRGATGRLLPEPGGFRHGTTHRRGRHGLPVGKCOA-----PGQSYDLC 120
DB 74 DAG-----DLDIEDMKRQNMCFYLKALDGFVWVLT 105
OY 121 SSSLIH-----NPTPG-TNFSLELIGHSIDFTHPCDOEELQDALTPRPNNLSKKKLEAPT 174
DB 106 DGDMIYISDNVKNKYMGLTF--ELTGHSVFEDTHPCDHEEMREMLTHRNGLVKKRGEONT 163
OY 175 ERHSLMKSTLTSRGTLNKAATWVVLHCSGHMRAKPPAOTSPPAGSPRSEPPLOCLV 234
DB 164 QRSFELMKCTLTSRGTMNKATWVVLHCTGHIHYDINSNPQCG--YKKPPTMCLV 221
OY 235 LICEAIP 241
DB 222 LICEPIR 228

RESULT 2
US-08-480-473b-2
; Sequence 2, Application US/08480473B
; Patent No. 5882914
; GENERAL INFORMATION:
; APPLICANT: Semenza, Gregg L.
; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,473B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-473b-2

Query Match 30.3%; Score 495; DB 2; Length 826;
Best Local Similarity 46.2%; Pred. No. 1.le-45;
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
```

```
OY 175 ERHSLMKSTLTSRGTLNKAATWVVLHCSGHMRAKPPAOTSPPAGSPRSEPPLOCLV 234
DB 164 QRSFELMKCTLTSRGTMNKATWVVLHCTGHIHYDINSNPQCG--YKKPPTMCLV 221
OY 235 LICEAIP 241
DB 222 LICEPIR 228

RESULT 3
US-08-915-213-2
; Sequence 2, Application US/08915213
; Patent No. 6020462
; GENERAL INFORMATION:
; APPLICANT: Semenza, Gregg L.
; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,213
; FILING DATE: 20-AUG-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/480,473
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-213-2

Query Match 30.3%; Score 495; DB 3; Length 826;
Best Local Similarity 46.2%; Pred. No. 1.le-45;
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
```



QY	235	LICEAIP	241
Db	222	LICEPIP	228

```

RESULT      4
US-09-148-547-2
; Sequence 2, Application US/09148547
; Patent No. 6124131
; GENERAL INFORMATION:
; APPLICANT: Semenza, Gregg L.
; TITLE OF INVENTION: Hypoxia Inducible Factor-1 and Methods of Use
; FILE REFERENCE: 07265/151001
; CURRENT APPLICATION NUMBER: US/09/148,547
; CURRENT FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 826
; TYPE: PR1
; ORGANISM: Homo sapiens
; US-09-148-547-2

```

RESULT 5  
US-09-235-217-2  
Sequence 2, Application US/09235217  
Patent No. 6222018  
GENERAL INFORMATION:  
APPLICANT: Semenza, Gregg L.  
TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/235,217  
FILING DATE:  
CLASSIFICATION:

```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER: US 08/480,473
3      FILING DATE:
4      ATTORNEY/AGENT INFORMATION:
5      NAME: Halle, Lisa A.
6      REGISTRATION NUMBER: 38,347
7      REFERENCE/DOCKET NUMBER: 07265/053001
8      TELECOMMUNICATION INFORMATION:
9      TELEPHONE: 619/678-5070
10     TELEFAX: 619/678-5099
11     INFORMATION FOR SEQ ID NO: 2:
12     SEQUENCE CHARACTERISTICS:
13     LENGTH: 826 amino acids
14     TYPE: amino acid
15     TOPOLOGY: linear
16     MOLECULE TYPE: protein
17     US-09-235-217-2

```

RESULT 6  
 PCT-US96-10251-2  
 Sequence 2, Application PC/TUS9610251  
 GENERAL INFORMATION:  
 APPLICANT: The Johns Hopkins University School of Medicine  
 TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
 NUMBER OF SEQUENCES: 35  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 4225 Executive Square, Suite 1400  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: USA  
 Zip: 92037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US96/10251  
 FILING DATE: 06-JUN-1996  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Halle, Lisa A.  
 REGISTRATION NUMBER: 38,347  
 REFERENCE/DOCKET NUMBER: 07265/053W01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619/678-5070

TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 826 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-10251-2

Query Match 30.3%; Score 495; DB 5; Length 826;  
Best Local Similarity 46.28; Pred. No. 1,1e-45;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;

QY 11 NTELKRSRDAARSRRSQTETVLYOLAHPLPARGVSAHLDRKASIMRLTISYLRMLIC 70  
DB 14 SSERKRSRDAARSRRKSESEVEFEIHAQLPLPHNVSSHLDKASVMLTISYLRVRL 73  
QY 71 AAGKRGKATGRLLPEPGGFRHGTTHRRGRGLPVGKCOQA-----PGPSVDLC 120  
DB 74 DAG-----DLIEDMKAKMNCFYLKALDGFVWLTD 105  
QY 121 SSLLH-----NPRPG-TNFSLELGHSTFDIHPDDEIADALTPRPNLSKKLEAPT 174  
DB 106 DGMVYISDNVKKYMGTLQF--ELGHSVDFTHPCDHEEKREMLTHRNGLVKRGKEQT 163  
QY 175 ERHFSLRMKSTLTSRGTINLKAATWVLCGSHMRAYKPPAOTSPAGSPRSEPPLOC 234  
DB 164 QRSFRLRMKCTLTSRGTINLKAATWVLCGSHMRAYKPPAOTSPAGSPRSEPPLOC 221  
QY 235 LVLCALP 241  
DB 222 LVLCALP 228

RESULT 7  
US-08-785-241-7  
Sequence 7, Application US/08785241  
Patent No. 5695963  
GENERAL INFORMATION:  
APPLICANT: McKnight, Steven L.  
APPLICANT: Russell, David W.  
TITLE OF INVENTION: Endothelial PAS Domain Protein  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 268 BUSH STREET, SUITE 3200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,241  
FILING DATE: 17-JAN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A.  
REGISTRATION NUMBER: 36,627  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 810 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-785-241-7

Query Match 29.9%; Score 489; DB 1; Length 810;  
Best Local Similarity 44.6%; Pred. No. 4.8e-45;  
Matches 111; Conservative 27; Mismatches 59; Indels 52; Gaps 5;

QY 11 NTELKRSRDAARSRRSQTETVLYOLAHPLPARGVSAHLDRKASIMRLTISYLRMLIC 70  
DB 2 SSERKRSRDAARSRRKSESEVEFEIHAQLPLPHNVSSHLDKASVMLTISYLRVRL 61  
QY 71 AAGKRGKATGR-----LLPEPGGFRHGTTHRRGRGLPVGKCOQAAG 113  
DB 62 DAGLDSEDKAKMDCFYLKALDGFVWLTDG----- 95  
QY 114 POSVDLCSSLIHNPTRG-TNFSLELGHSTFDIHPDDEIADALTPRPNLSKKLEA 172  
DB 96 ---DMVYISDNVKKYMGTLQF--ELGHSVDFTHPCDHEEKREMLTHRNGPVKRGKE 149  
QY 173 PTERHFSLRMKSTLTSRGTINLKAATWVLCGSHMRAYKPPAOTSPAGSPRSEPPLOC 232  
DB 150 NTOFSFLRMKCTLTSRGTINLKAATWVLCGSHMRAYKPPAOTSPAGSPRSEPPLOC 207  
QY 233 LVLCALP 241  
DB 208 LVLCALP 216

RESULT 8  
US-08-785-241-5  
Sequence 5, Application US/08785241  
Patent No. 5695963  
GENERAL INFORMATION:  
APPLICANT: McKnight, Steven L.  
APPLICANT: Russell, David W.  
TITLE OF INVENTION: Endothelial PAS Domain Protein  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 268 BUSH STREET, SUITE 3200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,241  
FILING DATE: 17-JAN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A.  
REGISTRATION NUMBER: 36,627  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 875 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-785-241-5

Query Match 29.2%; Score 477.5; DB 1; Length 875;  
Best Local Similarity 45.3%; Pred. No. 1e-43; Indels 27; Gaps 9;  
Matches 115; Conservative 33; Mismatches 79; Indels 27; Gaps 9;

QY 1 MALGIRVNSNTELRKESDAAARSRSQETEVLYQLAHTLPARGVSAHLDAKASIMRLT 60  
DB 1 MADAKKKRSSSELRKESDAAARCRSKETEVFYELAHELPLPHVSSHLDAKASIMRLA 60  
QY 61 ISYLRMHL-----CAAGKRGATGRLLPEGPGFRHGRHGRHGLPVGKCOQAPGPOS 116  
DB 61 ISFLFRHKLSSVCSNESAADQOM-----DNLTKALEGF-----IAVYTD 105  
QY 117 VLCSSSLHNPPTGNSFSLDELGHIFDFIHPDDELDALTPR--PULSKKLEAP 174  
DB 106 GMIPLFSENISKFMGLT-OVELTGHSHIFDTHPCDHEIRENLTKNGSGFGKSKDVST 164  
QY 175 ERHFSLRMSTLTSRGTNLKATWK-VLHCSGHRAYKP-PQTSPPAGSPSEPPLQC 232  
DB 165 ERDFEHRMCTVYNNRGTYNLKSATYKSVLHCTGVRYVNNCPHNSLCSG--KEPLSLC 222

QY 233 VLICEAIPOLPFH 246  
DB 223 LIMCEPI-QHPSH 235

RESULT 9  
US-08-785-241-4  
; Sequence 4, Application US/08785241  
; Patent No. 5695963  
; GENERAL INFORMATION:  
; APPLICANT: McKnight, Steven L.  
; APPLICANT: Russell, David W.  
; APPLICANT: Tian, Hui  
; TITLE OF INVENTION: Endothelial PAS Domain Protein  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785,241  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A.  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UTSD:1229  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4341  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 870 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-785-241-4

Query Match 28.5%; Score 466; DB 1; Length 870;  
Best Local Similarity 43.9%; Pred. No. 1.8e-42;  
Matches 111; Conservative 34; Mismatches 82; Indels 26; Gaps 8;

QY 1 MALGIRVNSNTELRKESDAAARSRSQETEVLYQLAHTLPARGVSAHLDAKASIMRLT 60

DB 1 MADAKKKRSSSELRKESDAAARCRSKETEVFYELAHELPLPHVSSHLDAKASIMRL 60  
QY 61 ISYLRMHL-----CAAGKRGATGRLLPEGPGFRHGRHGRHGLPVGKCOQAPGPOS 116  
DB 61 ISFLFRHKLSSVCSNESAADQOM-----DNLTKALEGF-----IAVYTD 105  
QY 117 VLCSSSLHNPPTGNSFSLDELGHIFDFIHPDDELDALTPR--PULSKKLEAP 174  
DB 106 GMIPLFSENISKFMGLT-OVELTGHSHIFDTHPCDHEIRENLTKNGSGFGKSKDMST 164  
QY 175 ERHFSLRMSTLTSRGTNLKATWK-VLHCSGHRAYKP-PQTSPPAGSPSEPPLQC 233  
DB 165 ERDFEHRMCTVYNNRGTYNLKSATYKSVLHCTGVRYVNNCPHNSLCSG--KEPLSLC 222

QY 234 VLICEAIPOLPFH 246  
DB 223 LIMCEPI-QHPSH 234

RESULT 10  
US-08-480-473B-3  
; Sequence 3, Application US/08480473B  
; Patent No. 5882914  
; GENERAL INFORMATION:  
; APPLICANT: Semenza, Gregg L.  
; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,473B  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hallie, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07265/053001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 373 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-480-473B-3

Query Match 27.1%; Score 444; DB 2; Length 373;  
Best Local Similarity 44.3%; Pred. No. 1.4e-40;  
Matches 104; Conservative 24; Mismatches 59; Indels 48; Gaps 6;

QY 23 ARSRQSETEVLYQLAHTLPARGVSAHLDAKASIMRLTISYLRMHLCAAGKRGATGR 82  
DB 5 AGSRKSESEVLYELAQLPLPHVSSHLDAKASIMRLTISYLRMHLCAAGKRGATGR 55  
QY 83 LPEGPGFRHGRHGRHGLPVGKCOQA-----PGQSVDCSSSLH-----N 127  
DB 56 -----DLIDIEDMKAOQNCFYALKALDGFVWVLLDDGDMITISDNVN 96

[illegible]

RESULT 11  
US-08-915-213-3  
: Sequence 3, Application US/08915213

1 TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
2  
3 NUMBER OF SEQUENCES: 64  
4  
5 CORRESPONDENCE ADDRESS:  
6 ADDRESSEE: Fish & Richardson P.C.  
7 STREET: 4225 Executive Square, Suite 1400  
8 CITY: La Jolla  
9 STATE: CA  
10 COUNTRY: USA  
11 ZIP: 92037  
12  
13 COMPUTER READABLE FORM:  
14 MEDIUM TYPE: Floppy disk  
15  
16 COMPUTER: IBM PC compatible  
17 OPERATING SYSTEM: PC-DOS/MS-DOS  
18 SOFTWARE: PatentIn Release #1.0, Version #1.30  
19  
20 CURRENT APPLICATION DATA:  
21 APPLICATION NUMBER: US/08/915,213  
22 FILING DATE: 20-AUG-1997  
23  
24 CLASSIFICATION: 514  
25  
26 PRIOR APPLICATION DATA:  
27 APPLICATION NUMBER: US 08/480,473  
28 FILING DATE: 06-JUN-1995  
29  
30 ATTORNEY/AGENT INFORMATION:  
31 NAME: Halle, Lisa A.  
32 REGISTRATION NUMBER: 38,347  
33 REFERENCE/DOCKET NUMBER: 07265/053001  
34 TELECOMMUNICATION INFORMATION:  
35 TELEPHONE: 619/678-5070  
36 TELEFAX: 619/678-5099  
37  
38 INFORMATION FOR SEQ ID NO: 3:  
39 SEQUENCE CHARACTERISTICS:  
40 LENGTH: 373 amino acids  
41 TYPE: amino acid  
42 STRANDEDNESS: not relevant  
43 TOPOLOGY: linear  
44  
45 MOLECULE TYPE: protein  
46  
47 US-08-915-213-3

Query Match	27.18	Score 444	DB 3	Length 373
Best Local Similarity	44.33	Pred. No. 1.4e+0		
Matches 104	Conservative 24	Mismatches 59	Indels 48	Gaps
QY	23	ARSRSQETEVLYOLAHNTLPKRCVSADLKASIMRLTISYLRMRKLCAGGKRGRATGR	82	
		:		
Db	5	AGSRSKSEVFYELAHQDLPRLPHVNSHLDKASVMRLTISLRKRLDLAG	55	
QY	83	LLPEPGGFRHGTHRGHGLPVGKCOQA-----PGQSVDLCSSSLIH-----N	127	
		:		
Db	56	-----DDIDEDMKAGQNCFFYLKALDGFVWVLDDGDMIYISDNVN	96	
QY	128	PTPG--TNPSELIIGHSTFDLPHPDQDELDALTPRRLSKKKEAETERHSIRMKSTL	186	
Db	97	KYMLTQF--ELTSHSVDFTHPDCDHMEREMLLHNRGLVKKGEQNTORSFFELRMKTL	154	
QY	187	TSRGRTNLKAAATKVVHCSGHMAAYPRPQTSFAGSRSPRLQCVILICEAP	241	
		:		
Db	155	TSRGRTNINISATKVVHCTGHINHYVDTPNSNQPGC--YKRPPTCVLICEPIP	207	

```

12 RESULT
13 US-09-235-217-3
14 : Sequence 3, Application US/09235217
15 : Patent No. 6222018
16 :
17 : GENERAL INFORMATION:
18 : APPLICANT: Semenza, Gregg L.
19 : TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
20 : NUMBER OF SEQUENCES: 64
21 : CORRESPONDENCE ADDRESS:
22 : ADDRESSEE: Fish & Richardson P.C.
23 : STREET: 4225 Executive Square, Suite 1400
24 : CITY: La Jolla
25 : STATE: CA
26 : COUNTRY: USA
27 :
28 : ZIP: 92037
29 : COMPUTER READABLE FORM:
30 : MEDIUM TYPE: Floppy disk
31 : COMPUTER: IBM PC compatible
32 : OPERATING SYSTEM: PC-DOS/MS-DOS
33 : SOFTWARE: PatentIn Release #1.0, Version #1.30
34 :
35 : CURRENT APPLICATION DATA:
36 : APPLICATION NUMBER: US/09/235,217
37 : FILING DATE:
38 : CLASSIFICATION:
39 : PRIOR APPLICATION DATA:
40 : APPLICATION NUMBER: US 08/480,473
41 : FILING DATE:
42 :
43 : ATTORNEY/AGENT INFORMATION:
44 : NAME: Halle, Lisa A.
45 : REGISTRATION NUMBER: 38,347
46 : REFERENCE/DOCKET NUMBER: 07265/053001
47 : TELECOMMUNICATION INFORMATION:
48 : TELEPHONE: 619/678-5070
49 : TELEFAX: 619/678-5099
50 :
51 : INFORMATION FOR SEQ ID NO: 3:
52 : SEQUENCE CHARACTERISTICS:
53 : LENGTH: 373 amino acids
54 : TYPE: amino acid
55 : STRANDEDNESS: not relevant
56 : TOPOLOGY: linear
57 :
58 : MOLECULE TYPE: protein
59 :
60 : US-09-235-217-3

```

```

Query Match      27.1%  Score 444;  DB 4;  Length 373;
Best Local Similarity 44.3%;  Pred. No. 1.4e-40;
Matches 104;  Conservative 24;  Mismatches 59;  Indels 48;  Gaps 6;

QY      23  ARSRQETEVLYQLAHTLPFARGVSAHLMDKASIMRLTISYRMHRLCAAGKRGKATGR  82
      | | | | | : | | | | | : | | | | | : | | | | | : | |
Db      5  AGSRKSKSEVEYELAHQLPLPHNVSSHLMDKASVRLTISYLRVKKLDAG-----  55

QY      83  LLPEGGGFRHGTNRGRHGLPVGKCOA-----PGQSYDLCSSLIH-----N  127
      | : : : : : | : : : : : | : : : : : | : : : : : | :
Db      56  -----DLDIEDMKAKQNCFLYALKALDGFVMVLTDDGDMIYISDNNV  96

QY      128  PTPG-TNSLEIGHISIDFTHPCOEELQDALTRPNISSKKLAPTRHNSLNRKSTL  186
      | | | | | : | | | | | : | | | | | : | | | | | : | |
Db      97  KYMGTLQF-EITGSHVEDFTTHPCDHEEREMLTFRNGVLGKKGKEQNTORSEFLRMKTL  154

QY      187  TSRGTLMLKATWVVLHCSGHMRAVYKRPAGQSPGASPRSEPPLOCIVLICEAIP  241
      | | | | | : | | | | | : | : : : : | : : : : : | :
Db      155  TSGRTNMKISATWVVLHCTGHIHYIDNSNPQCG--YKRPMTCLVLICEPIP  207

RESULT  13
PCT-0596-10251-3
; Sequence 3, Application PC/RTUS9610251
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University School of Medicine
; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
; NUMBER OF SEQUENCES: 35

```

```

      SOFTWARE: Patentin Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/480,473B
      FILING DATE: 06-JUN-1995
      CLASSIFICATION: 514
      ATTORNEY/AGENT INFORMATION:
      NAME: Halle, Lisa A.
      REGISTRATION NUMBER: 38,347
      REFERENCE/DOCKET NUMBER: 07265/053001
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 619/678-5070
      TELEFAX: 619/678-5099
      INFORMATION FOR SEQ ID NO: 4:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 805 amino acids
      TYPE: amino acid
      STRANDEDNESS: not relevant
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      US-08-480-473B-4

Query Match      27.1%; Score 444; DB 2; Length 805;
Best Local Similarity 44.3%; Pred. NO. 4.4e-40;
Matches 104; Conservative 24; Mismatches 59; Indels 48; Gaps 6.

      23 ARSRSOETVELLYOAHRLPFARGVSAHLDRKASIMRLTITSYRMHRLCAAGSKRGATGR 82
      | |||||::|||::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      Db 5 AGSRRSKSEVFYELAHOLDPLPHNVSSHLDRKASVRLTISYLRVKLLDAG----- 55
      83 LLPEGGEGFRGHRHGRHGLPVGKCOQA-----PGQSVYDLCSSSLIH-----N 127

```

0Y 128 PTPB-TNFSLELLGHSTFDFTHPCQDEELDALTPRPNLSKKKLEAPRHRHSLRMKSTL 186  
 Db 97 KYMGLTOR--FLTGSVDFETHPCHEMRREXLTJHRNGLVKRGKQONQORSFLLRMKCTL 154  
 0Y 187 TSRGRTNLKAAATMKVNLHCSGHMRAYPKPPACTSPAGSPRSEBPLLOCLVYICAI 241  
 155 TSRGRTNMRKATMKVNLHCTGHIHYDINSNPQCG--YKKRPMICVLICEPIP 207  
 RESULT 15  
 US-08-915-213-4  
 ; Sequence 4, Application US/08915213  
 ; Patent No. 6020462  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Semenza, Gregg L.  
 ; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
 ; NUMBER OF SEQUENCES: 64  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 4225 Executive Square, Suite 1400  
 ; City: La Jolla  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/915,213  
 ; FILING DATE: 20-AUG-1997  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/480,473  
 ; FILING DATE: 06-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Halle, Lisa A.  
 ; REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 07265/053001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619/678-5070  
 TELEFAX: 619/678-5099  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 805 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-915-213-4

Query Match 27.1%; Score 444; DB 3; Length 805;  
 Best Local Similarity 44.3%; Pred. No. 4.4e-40;  
 Matches 104; Conservative 24; Mismatches 59; Indels 48; Gaps 6;

QY	23	AKSRRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHLCAAGGKRGRTGR	82
Db	5	AGSRRSKSESEVFEYLAHOLPLPHNVSSHLDKASVRLTISYLRVRLDAG-----	55
QY	83	LLPEGPGGFRHGTNRGRHGLPYGKCOA-----PGQSYDLCSSSLIH-----N	127
Db	56	-----DIDIEDMKRQNMCFYKALDGFVAVLTDDGDMYISDNVN	96
QY	128	PTPG-TNFSLELIGHIFDFTHPCDQELDADLPFPNLSKKKLEAPTERHFSLRKSTL	186
Db	97	KYMGITGF-ELTGHVSDFTHPCDHEMRMLTHRNLVKKGRQNTORSFFLRKCTL	154
QY	187	TSRGFTLTKAATWKVLCGSHMRAYKPPAOTSPPAGSPRSEPPLOCLVLICEAIP	241
Db	155	TSRGFTYMKASATWKVLCGTHIHVYDTNSNQPCG--YKKPPTCLVLICEPIP	207

Search completed: September 8, 2002, 07:33:59  
 Job time: 2907 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2002, 07:32:32 ; Search time 37.72 Seconds  
(without alignments)  
782.063 Million cell updates/sec

```

Title: US-09-896-791B-3
Perfect score: 1636
Sequence: 1 MALGLQRYRVSNTLELRKESR.....TESSLPSPWLMALNRKNCPG 307

```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

Database :      PIR_71:*
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*

A:Title: Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in
A:Reference number: JC7771; PMID:11573933
A:Contents: Kidney
A:Accession: JC7771
A:Molecule type: mRNA

```

<p>Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.</p> <p>A:Cross-references: DDBJ:AB054067</p> <p>C:Comment: This protein is a heterodimeric transcription factor that belongs to the b</p> <p>Ived in the regulation of hypoxia-inducible gene expression in human kidney.</p>	
---	--

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	802	49.0	667	2	JC7771	hypoxia inducible
2	503	30.7	811	2	JC7619	hypoxia-inducible
3	495	30.3	826	2	I38972	hypoxia-inducible
4	492	30.1	810	2	JC4837	hypoxia-inducible
5	485	29.6	813	2	JC5809	hypoxia-inducible
6	281	17.2	1505	2	JC4851	hypoxia-inducible
7	256	15.6	248	2	A58520	single-minded gene
8	202	12.3	655	2	A29945	neurogenesis regul
9	169.5	10.4	823	2	T21943	neurogenesis regul
10	169.5	10.4	825	2	T21944	hypothetical prot
11	142	8.7	805	2	JC7635	hypothetical prot
12	138.5	8.5	392	2	JC7633	aryl hydrocarbon r
13	138.5	8.5	791	2	A56241	aryl hydrocarbon r
14	136.5	8.3	776	2	A55448	aryl hydrocarbon r
15	136.5	8.3	789	2	I55950	Ah receptor nucle
16	129.5	7.9	358	2	S58376	aryl hydrocarbon r
17	129.5	7.9	451	2	T42387	aryl hydrocarbon r
18	129	7.9	453	2	T19440	aryl hydrocarbon r
19	128	7.8	36	2	A14647	hypothetical prot
20	125.5	7.7	626	2	JC5405	neurogenesis regul
21	123.5	7.5	626	2	JE0270	brain and muscle /
22	106	6.5	848	2	S59514	Ant1-like PAS prok
23	104.5	6.4	1059	2	T30557	aryl hydrocarbon r
24	103.5	6.3	258	2	JC5407	aryl hydrocarbon r
25	103	6.3	846	2	JC7721	brain and muscle /
26	103	6.3	1079	2	A70972	aryl hydrocarbon r
27	102.5	6.3	271	2	PC4288	probable DNA poly
28	102	6.2	716	2	T09462	brain and muscle /
29	98	6.0	1058	2	T30556	juvenile hormone r
						aryl hydrocarbon r

30	97.5	6.0	326	2	\$71755	transcription factor
31	97	5.9	527	2	\$75399	hypothetical protein
32	95.5	5.8	920	2	JC7313	aryl hydrocarbon r
33	94.5	5.8	300	2	T24292	hypothetical proteol
34	94.5	5.8	1151	2	T24541	hypothetical proteol
35	93.5	5.7	634	2	T00359	hypothetical prote
36	92.5	5.7	1751	2	T09394	hypothetical prote
37	90	5.5	743	2	G83726	gag-pro-pol polypr
38	90	5.5	1920	2	T13893	assimilatory nitrat
39	89.5	5.5	1891	2	T13594	gene hindsight pro
40	89	5.4	853	2	S58375	hypothetical prote
41	89	5.4	3133	2	S52093	aryl hydrocarbon r
42	87.5	5.3	647	2	S06450	hemocytin - silkw
43	87.5	5.3	805	2	A46265	steroid hormone re
44	87.5	5.3	2481	2	A43908	aryl hydrocarbon r
45	87	5.3	1140	2	B70729	flavonectin - Afr
						hypothetical prote

## ALIGNMENTS

RESULT 1  
JC7771  
hypoxia inducible factor-3 alpha - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: JC7771  
R:Hara, S.; Hamada, J.; Kobayashi, C.; Kondo, Y.; Imura, N.  
Biochem. Biophys. Res. Commun. 287, 808-813, 2001  
A:Title: Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in  
A:Reference number: JC7771; PMID:11573933  
A:Contents: Kidney  
A:Accession: JC7771  
A:Molecule type: mRNA  
A:Residues: 1-667 <HAR>  
A:Cross-references: DDBJ:AB054067  
C:Comment: This protein is a heterodimeric transcription factor that belongs to the  
lived in the regulation of hypoxia-inducible gene expression in human kidney.  
C:Genetics:  
A:Gene: hlf-3alpha  
A:Map position: 19  
C:Keywords: kidney

Query Match	49.0%	Score 802	DB 2	Length 667
Best Local Similarity	68.0%	Pred. No. 7.4e-62		
Matches 172	Conservative 11	Mismatches 28	Indels 42	Gaps 5
QY	9	RNTETLRKRSKSDAARSRROETEVYLYOAHPLPFRAGVSAHLDXASIMRLTISYLRMR	68	
Db	7	RSTTELRRKRSKSDAARSRROETEVYLYOAHPLPFRAGVSAHLDXASIMRLTISYLRMR	66	
QY	69	LCAAGGKRGARAGLLPERRPGGFRHGTNRGRHGLPVGCKQ-----QAPGPOSV	117	
Db	67	LCAAG-----EWNVYGAGEPLDACYKALKEGVWVLTABSDMAY	106	
QY	118	DLCSSSLINHPPTGTFNS-TELIGHISIFPIPCDDEELDALTTPPNLSKKKLEAPTER	176	
Db	107	-----LSENVSKHGLSOLLEIGHISIFPIPCDDEELDALTTPQUTLSKKKLEAPTER	160	
QY	177	HPSLNRKSTLSRGRITLNKAATWKVLHSCGIMRAVKKPPAQTSPAGSPSEPPLOCLVLI	236	
Db	161	CPSLNRKSTLSRGRITLNKAATWKVLKNSGIMRAVKKPPAQTSPAGSPSEPPLOCLVLI	220	
QY	237	CEAIPDLPEHDGA	249	
Db	221	CEAIP-----HPSG	229	
RESULT	2			
JC7619				
hypoxia-inducible factor 1 alpha - chicken				
C:Species: Gallus gallus (chicken)				







```

Db      60 LCDAMQ--PSRAGPLDGVAKELGSHLLQTDGFVFFVASDGMKIMYISETASVHLGLS-- 115
QY      125 INNPRTGTFNPSLELGHSTFDFIHPDQDELDALTPRPNLSKKKL-EAPTEHSELRMK 183
Db      116 -----QVELTGNSTIYEYTHPSDHDMATVLAHQPLHHLLQEEYEIERSEFLRMK 165
QY      184 STLTSGRTLNKAAATWKVLHCSGHRM 210
Db      166 CVLAK--RNAGLTCSGKYIHCSGYLK 190

RESULT  8
A29945
neurogenesis regulatory protein - fruit fly (Drosophila melanogaster) (fragment)
N:Alternate names: single-minded gene protein
C:Species: Drosophila melanogaster
C>Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Mar-1998
C:Accession: A29945
R:Crewe, S.T.; Thomas, J.B.; Goodman, C.S.
Cell 52, 143-151, 1988
A>Title: The Drosophila single-minded gene encodes a nuclear protein with sequence siml
A:Reference number: A29945; MUID:88151023
A:Accession: A29945
A:Molecule type: mRNA
A:Residues: 1-655 <CRE>
A:Cross-references: GB:M19020; NID:q158464; PID:q158465
C:Genetics:
A:Gene: slm
A:Cross-references: FlyBase:FBgn0004666
C:Keywords: DNA binding; transcription regulation

Query Match      12.3%; Score 202; DB 2; Length 655;
Best Local Similarity 28.5%; Pred. No. 1.3e-09;
Matches 73; Conservative 35; Mismatches 78; Indels 70; Gaps 10;

QY      36 QLAHTLPFARGVSAHLDKASIMRLTISYLRMRLCAAGKRGKRGATGRLLPEGPGGFRHGT 95
Db      4 ELAKLLPLPAATISQLDKASVIRLTISYLRMKRVFPDG--LGRAMG-----SSPAMRGAT 57
QY      96 HRR-GRHG-----PVGKCOAPQPSVDLCSSSLHNPPTGTFNLSLELGHST 143
Db      58 IRELGSHELLQTDGFLTFVVAAPDGKIMYISETASVHLGLS-----QVELTGNST 105
QY      144 PFIHPDQDELDALTPRPNLSKKKL-----EAPT-----EER 177
Db      106 FEYIHHYDDEMAATISLHPHINQHPDLAQTTPPGSPGCVQHPASAYDHDGSHTEIEKT 165
QY      178 FSLRMKSTLTSGRTLNKAAATWKVLHCSGHRMAYKPPAQTSPAGSPRESEPLQCLVLIC 237
Db      166 FFLRMKCVLAK--RNAGLTSGFKYIHCSGYLKARITPPRGDQGS-----LIQNLGIVA 218
QY      238 -----EAIPLPFH 246
Db      219 VGHSLPSSAITEIKH 234

RESULT  9
T21943
hypothetical protein F38A6.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21943; T26899
R:Mortimore, B.
Submitted to the EMBL Data Library, March 1997
A:Reference number: Z19491
A:Accession: T21943
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-823 <M12>
A:Cross-references: EMBL:Z92833; PIDN:CA807380.1; GSPDB:GN00023; CESP:F38A6.3a
A:Experimental source: clone F38A6

```

```

R:Ainscough, R.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z20282
A:Accession: T26899
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-823 <M12>
A:Cross-references: EMBL:AL023842; PIDN:CAA19520.1; GSPDB:GN00023; CESP:F38A6.3a
A:Experimental source: clone Y44A6D
C:Genetics:
A:Gene: CESP:F38A6.3a
A:Map position: 5
A:introns: 19/2; 75/1; 111/2; 175/1; 226/3; 331/3; 518/1; 627/2; 711/1; 768/3

Query Match      10.4%; Score 169.5; DB 2; Length 823;
Best Local Similarity 26.3%; Pred. No. 1.2e-06;
Matches 62; Conservative 35; Mismatches 90; Indels 49; Gaps 10;

QY      9 RSNTELKREKSRDARSRSOETEVLYQLAHTLPFA-RGVSAHLDKASIMRLTISYLRMH 67
Db      110 KRMERKRETSRHAADRKSSEDIFDDLMKCVPIYEEGTVHHDLRALRVAATICRLR 169
QY      68 RLCA--AGCKRGATGRLLPEGPGGFRHGTGRHGLPVGKCOAPQPSVDLCSSSL 124
Db      170 KTAGVLENNLNETNEWTED-----TIAECLDG-FVMIVSDSSYL 212
QY      125 INNPRTGTFNPSL--ELIGHSTFDFIHPDQDELDALTPRPNLSKKKLEAP-----T 174
Db      213 YTESVAMYLGLQTDLTGRALRDLFLPSDYDFE-----DKOSKMLHKRPGEDTDTT 264
QY      175 ERHFSLRMKSTLTSGRTLNKAAATWK-----VLH-----CSGHRMAYKPPAQTSPAG 222
Db      265 GINMVRMKMTYISPRRCNLMSALYKSVFLVHKSIVSGHVSFMQ--GITIPAG 318

RESULT  10
T21944
hypothetical protein F38A6.3b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21944; T26900
R:Mortimore, B.
Submitted to the EMBL Data Library, March 1997
A:Reference number: Z19491
A:Accession: T21944
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-825 <M12>
A:Cross-references: EMBL:Z92833; PIDN:CA807381.1; GSPDB:GN00023; CESP:F38A6.3b
A:Experimental source: clone F38A6
R:Ainscough, R.
Submitted to the EMBL Data Library, June 1998
A:Reference number: Z20282
A:Accession: T26900
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-825 <M12>
A:Cross-references: EMBL:AL023842; PIDN:CAA19521.1; GSPDB:GN00023; CESP:F38A6.3b
A:Experimental source: clone Y44A6D
A:Gene: CESP:F38A6.3b
A:Map position: 5
A:introns: 19/2; 75/1; 111/2; 175/1; 226/3; 331/3; 518/1; 627/2; 713/1; 770/3

Query Match      10.4%; Score 169.5; DB 2; Length 825;
Best Local Similarity 26.3%; Pred. No. 1.2e-06;
Matches 62; Conservative 35; Mismatches 90; Indels 49; Gaps 10;

QY      9 RSNTELKREKSRDARSRSOETEVLYQLAHTLPFA-RGVSAHLDKASIMRLTISYLRMH 67
Db      110 KRMERKRETSRHAADRKSSEDIFDDLMKCVPIYEEGTVHHDLRALRVAATICRLR 169

```

OY 68 RLCA---AGGKRGATGRLLPEGPGFRHGRHGRHGLPVKCOQAPGQSDLCSSSL 124  
 A:Residues: 1-392 <HSD>  
 A:Cross-References: GB:AY007992  
 C:Comment: This factor, a basic helix-loop-helix PAS (bHLH-PAS) factor, heterodimeriz  
 C:Gene: arntl2x  
 DB 170 KTAGVLENNLDNETINTEWTD-----TIAECIDG-FWMTVDSDSL 212  
 C:Species: Gallus gallus (chicken)  
 C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 30-Jun-2001  
 C:Accession: J07635  
 R:Ca tron, T.; Mendioja, M.A.; Smith, S.M.; Born, J.; Walker, M.K.  
 Biochem. Biophys. Res. Commun. 282, 602-607, 2001  
 A:Title: Hypoxia regulates avian cardiac Arnt and HIF-1alpha mRNA expression.  
 A:Reference number: J07635; MUID:21294777; PMID:11401503  
 A:Contents: Embryo  
 A:Accession: J07635  
 A:Molecule type: mRNA  
 A:Residues: 1-805 <CAT>  
 C:Comment: This protein, known as hypoxia-inducible factor 1 beta, a member of the basic  
 C:Genetics: responses to hormonal and environmental stimuli. It is also a common dimer partner for tran  
 A:Gene: arntl  
 C:Keywords: transcription factor  
 RESULT 11  
 J07635  
 aryl hydrocarbon receptor nuclear translocator 1 - chicken  
 N:Alternate names: hypoxia-inducible factor 1 beta  
 C:Species: Gallus gallus (chicken)  
 C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 30-Jun-2001  
 C:Accession: J07635  
 R:Ca tron, T.; Mendioja, M.A.; Smith, S.M.; Born, J.; Walker, M.K.  
 Biochem. Biophys. Res. Commun. 282, 602-607, 2001  
 A:Title: Hypoxia regulates avian cardiac Arnt and HIF-1alpha mRNA expression.  
 A:Reference number: J07635; MUID:21294777; PMID:11401503  
 A:Contents: Embryo  
 A:Accession: J07635  
 A:Molecule type: mRNA  
 A:Residues: 1-805 <CAT>  
 C:Comment: This protein, known as hypoxia-inducible factor 1 beta, a member of the basic  
 C:Genetics: responses to hormonal and environmental stimuli. It is also a common dimer partner for tran  
 A:Gene: arntl  
 C:Keywords: transcription factor

Query Match 8.7%; Score 142; DB 2; Length 805;  
 Best Local Similarity 20.5%; Pred. No. 0.00028;  
 Matches 61; Conservative 43; Mismatches 95; Indels 98; Gaps 8;  
 OY 14 LRKESRDAARSRSOETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHLCAAG 73  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 89 LARENHSEIERRRRNKMTAYITELSDMVPYCSALARKPKLITLRMAVSHM----- 139  
 A:Accession: A56241  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 74 GKRGATGRLLPEGPGFRHGRHGRHGLPVKCOQAPGQSDLCSSSL-----HNPT 129  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 140 -KSLGTGNTSTDGKYKPSFLTDQELKHLI-----LEADGFLFYVSCETGRVYVSDSVT 194  
 C:Accession: A56241  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 130 PGTNF-STELIGHSTFDFIHPDOELDALTPRPV-LSKKLEAPTERHF 178  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 195 PVLNPOQSEWPGSTLYDQVHPRDVKLRQLSTSENALTEGKPKCLSKKDPAAAPES-- 252  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 179 SLRMKSTLTSRGRTLNLAATWK----- 201  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 253 -----ASKGRILDKTGYVKEGQSMRMCGSRSEFICRMCGSSVDVAVNRLS 304  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 202 -----VLHSGHMRAYKPPAQTSPAGSP-RSEPPLOCVLVI 236  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 305 FMRNRCRNGLTKDGEHYVYVHCTGYIKAMPAGVSLPDDDPAGQGSKECLVAI 361  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 RESULT 12  
 J07633  
 aryl hydrocarbon nuclear translocator ARNT2-like factor, ARNT2X - zebra fish  
 C:Species: Brachydanio rerio (zebra fish)  
 C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 30-Jun-2001  
 C:Accession: J07633  
 R:Hsu, H.J.; Wang, W.D.; Hu, C.H.  
 Biochem. Biophys. Res. Commun. 282, 487-492, 2001  
 A:Title: Ectopic expression of negative ARNT2 factor disrupts fish development.  
 A:Reference number: J07633; MUID:21294759; PMID:11401485  
 A:Contents: Heart  
 A:Accession: J07633

A:Molecule type: mRNA  
 A:Residues: 1-392 <HSD>  
 A:Cross-References: GB:AY007992  
 C:Comment: This factor, a basic helix-loop-helix PAS (bHLH-PAS) factor, heterodimeriz  
 C:Gene: arntl2x  
 Query Match 8.5%; Score 138.5; DB 2; Length 392;  
 Best Local Similarity 20.8%; Pred. No. 0.00025;  
 Matches 55; Conservative 48; Mismatches 103; Indels 59; Gaps 8;  
 OY 16 KEKSRDAARSRSOETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHLCAAGK 75  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 47 RENHSEIERRRRNKMTAYITELSDMVPYCSALARKPKLITLRMAVSHM-----K 96  
 A:Accession: A56241  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 76 RGRATGRLLPEGPGFRHGRHGRHGLPVKCOQAPGQSDLCSSSL-----HNPTG 131  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 97 SMRGNTSTDGAYKPSFLTDQELKHLI-----LEADGFLFYVAAETGRVYVSDSVTV 152  
 C:Accession: A56241  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 132 TNF-STELIGHSTFDFIHPDOELDALTPRPVLSKKLEAP----- 170  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 153 LNHPOSEWPGSTLYDQVHPRDVKLRQLSTSENALTEGKPKCLSKKDPAAAPES-- 212  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 171 -----EAPTRHFLRMKSTLTSRGRT-----LNLKATWVVLHSGHMRAY 212  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 213 MGRSFTICRMCGSAPLD-HISLNRSSMKRKYRNGLGRPSKEGAQSYVHCTGYIKAM 271  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 213 KPPAQTSP-AGSPREPPLOCVLVI 236  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 272 PPAQMTIPDEDTAGQTSKYCLVAI 296  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT

RESULT 13  
 A56241  
 aryl hydrocarbon receptor nuclear translocator protein - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 05-Nov-1999  
 C:Accession: A56241  
 R:Reisz-Porras, S.; Probst, M.R.; Fukunaga, B.N.; Hankinson, O.  
 Mol. Cell. Biol. 14, 6075-6086, 1994  
 A:Title: Identification of functional domains of the aryl hydrocarbon receptor nuclea  
 A:Reference number: A56241; MUID:94344118  
 A:Accession: A56241  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 Query Match 8.5%; Score 138.5; DB 2; Length 791;  
 Best Local Similarity 21.8%; Pred. No. 0.00056;  
 Matches 58; Conservative 50; Mismatches 101; Indels 57; Gaps 9;  
 OY 14 LRKESRDAARSRSOETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHLCAAG 73  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 89 LARENHSEIERRRRNKMTAYITELSDMVPYCSALARKPKLITLRMAVSHM----- 139  
 A:Accession: A56241  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 74 GKRGATGRLLPEGPGFRHGRHGRHGLPVKCOQAPGQSDLCSSSL-----HNPT 129  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 140 -KSLRGNTSTDGKYKPSFLTDQELKHLI-----LEADGFLFYVSCETGRVYVSDSVT 194  
 C:Accession: A56241  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 130 PGTNF-STELIGHSTFDFIHPDOELDALTPRPV-LSKKLEAPTERHF 174  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 195 PVLNPOQSEWPGSTLYDQVHPRDVKLRQLSTSENALTEGKPKCLSKKDPAAAPES-- 254  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 OY 175 -----ERHFSLRMKSTLTS-----RGRTLNLAAT-----MKVLHSGHMRAY 211  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT  
 DB 255 MCMGSRSFICRMCGSAPLD-HISLNRSSMKRKYRNGLGRPSKEGAQSYVHCTGYIKAM 314  
 A:Residues: 1-791 <REI>  
 A:Cross-References: GB:U10325; NID:9555687; PID:AAA5671.1; PID:9555688  
 A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
 C:Genetics: A:Note: authors failed to translate CAG for residue 507 as Gln  
 A:Gene: ARNT





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2002, 07:33:27 ; Search time 19.87 Seconds

(without alignments)  
598.234 Million cell updates/sec

Title: US-09-896-791b-3

Perfect score: 1636  
Sequence: 1 MALTGRVRSNTLRRKRSR.....TESSLPSVLMALNRKNCPC 307

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	495	30.3	822	1 HIFA_MOUSE	Q61221 mus musculu
2	495	30.3	826	1 HIFA_HUMAN	Q16665 homo sapien
3	488	29.8	874	1 PAS1_MOUSE	P97481 mus musculu
4	467	28.5	870	1 PAS1_HUMAN	O99814 homo sapien
5	294	18.0	590	1 NPAL_MOUSE	O99742 homo sapien
6	293.5	17.9	594	1 NPAL_HUMAN	P97459 mus musculu
7	281	17.2	1507	1 SIMA_MOUSE	Q24167 drosophila
8	268	16.4	765	1 SIMA_HUMAN	P81133 homo sapien
9	266	16.3	766	1 SIM1_MOUSE	P81045 mus musculu
10	265	16.2	657	1 SIM2_MOUSE	P81079 mus musculu
11	256	15.6	667	1 SIM2_HUMAN	Q14190 homo sapien
12	251	15.3	958	1 TRH_MOUSE	Q24119 drosophila
13	249	15.2	673	1 SIM_MOUSE	P05709 drosophila
14	149	9.1	413	1 CYCL_MOUSE	O61734 drosophila
15	138.5	8.5	791	1 ARNT_MOUSE	P53762 mus musculu
16	137.5	8.4	800	1 ARNT_HUMAN	P41739 ratu musculu
17	136.5	8.3	789	1 ARNT_MOUSE	P27540 homo sapien
18	135.5	8.3	790	1 ARNT_HUMAN	O02748 oryctolagus
19	131.5	8.0	706	1 ARNT_MOUSE	O9hb22 homo sapien
20	131.5	8.0	712	1 ARNT_HUMAN	O61324 mus musculu
21	126	7.7	644	1 ARNT_MOUSE	O15945 drosophila
22	122.5	7.5	583	1 BMAL_MOUSE	O00327 homo sapien
23	106	6.5	848	1 AHR_MOUSE	P35869 homo sapien
24	91.5	5.6	846	1 CLOC_MOUSE	O15116 homo sapien
25	91	5.6	3133	1 AHR_MOUSE	P98092 bombix mori
26	90.5	5.5	805	1 AHR_MOUSE	P30561 mus musculu
27	90.5	5.5	855	1 CLOC_MOUSE	O08785 mus musculu
28	89.5	5.5	304	1 HEY1_MOUSE	O9y513 homo sapien
29	89	5.4	853	1 AHR_MOUSE	P41738 ratu musculu
30	87.5	5.3	471	1 RTN2_MOUSE	O97022 mus musculu
31	87.5	5.3	587	1 RTN2_HUMAN	O97023 mus musculu
32	87.5	5.3	602	1 ELL_MOUSE	O08856 mus musculu
33	87.5	5.3	647	1 KNR1_MOUSE	P13054 drosophila

34	87.5	5.3	2481	1 FINC_XENLA	Q91740 xenopus lae
35	86.5	5.3	545	1 RTN2_HUMAN	O75298 homo sapien
36	86	5.3	700	1 MRL1_CHICK	O91am7 gallus gall
37	85	5.2	475	1 SNA2_MOUSE	O62203 mus musculu
38	84	5.1	213	1 NUTL_MOUSE	P32845 drosophila
39	83	5.1	304	1 HEY1_MOUSE	O91am7 gallus gall
40	83	5.1	426	1 SNA7_HUMAN	O15105 homo sapien
41	83	5.1	1024	1 RIP3_MOUSE	P97434 mus musculu
42	82.5	5.0	649	1 V7OK_MOUSE	P20129 eggplant mo
43	82.5	5.0	711	1 MRL1_XENLA	O9w6k1 xenopus lae
44	82.5	5.0	1159	1 HERG_HUMAN	O12809 homo sapien
45	82.5	5.0	1487	1 ICP4_HSVB	P28925 equine herp

#### ALIGNMENTS

RESULT 1  
ID HIFA\_MOUSE STANDARD; PRT; 822 AA.  
AC 061221: 061665: 061664: 008993: 008741;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (ARNT interacting protein).  
CN HIFA.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Hepatocytes;  
RX MEDLINE=96353491; PubMed=8702901;  
RA Li H., Ko H.P., Whitlock J.P. Jr.;  
RT "Induction of phosphoglycerate kinase 1 gene expression by hypoxia.  
RT Roles of Arnt and Hif1alpha.";  
RT J. Biol. Chem. 271:21262-21267(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SVJ;  
RX MEDLINE=98034461; PubMed=9368100;  
RA Luo G., Gu Y.-Z., Jain S., Chan W.K., Carr K.M., Hogenesch J.B.,  
RA Bradfield C.A.;  
RT "Molecular characterization of the murine Hif-1 alpha locus.";  
RL Gene Expr. 6:287-299(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SVJ;  
RX MEDLINE=97354184; PubMed=9210478;  
RA Wenger R.H., Rolfs A., Kvietikova I., Spielmann P., Zimmermann D.R.,  
RA Gassmann M.;  
RT "The mouse gene for hypoxia-inducible factor-1alpha. Genomic organization, expression and characterization of an alternative first exon and 5' flanking sequence.";  
RT Eur. J. Biochem. 246:155-165(1997).  
RN [4]  
RP SEQUENCE OF 13-822 FROM N.A.  
RC TISSUE=Hepatocytes;  
RX MEDLINE=96254028; PubMed=8660378;  
RA Wenger R.H., Rolfs A., Marti H.H., Guenet J.-L., Gassmann M.;  
RT "Nucleotide sequence, chromosomal assignment and mRNA expression of mouse hypoxia-inducible factor-1 alpha.";  
RL Biochem. Biophys. Res. Commun. 223:54-59(1996).  
RC TISSUE=Hepatocytes;  
RA O'Rourke J.F.;  
RP SEQUENCE OF 22-85 FROM N.A.  
RT Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: INVOLVED IN THE INDUCTION OF OXYGEN REGULATED GENES.  
CC SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE ELEMENT (HRE).  
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER

CC BHLH PROTEIN. HETERODIMER OF AN ALPHA AND A BETA (ARNT) SUBUNITS.  
CC THE ALPHA SUBUNIT IS UNIQUE TO HIF-1 WHEREAS HIF-1 BETA (ARNT) CAN  
CC DIMERIZE WITH OTHER BHLH-PAS PROTEINS. INTERACTS WITH HSP90 (BY  
CC SIMILARITY)..  
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).  
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.  
CC -1- DOMAIN: BOTH HYPOXIA RESPONSIVENESS AND TRANSACTIVATION CAPABILITY  
CC RESIDE WITHIN THE C-TERMINAL PART.  
CC -1- PM: REQUIRES PHOSPHORYLATION FOR DNA-BINDING.  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS.  
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL:	U59496:	AAC52730.1:	-	
DR	EMBL:	AF003695:	AAC53455.1:	-	
DR	EMBL:	Y13685:	CAA70306.1:		
DR	EMBL:	Y13655:	CAA70306.1:	JOINED.	
DR	EMBL:	Y09085:	CAA70305.1:	-	
DR	EMBL:	AF004155:	AAC53461.1:	-	
DR	EMBL:	AF004141:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004142:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004143:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004144:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004145:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004146:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004147:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004148:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004149:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004150:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004151:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004152:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004153:	AAC53461.1:	JOINED.	
DR	EMBL:	AF004154:	AAC53461.1:	JOINED.	
DR	EMBL:	X95580:	CAA64833.1:	-	
DR	EMBL:	X95002:	CAA64458.1:	-	
DR	MGI:	MGI:106918:	H4f1a.		
DR	InterPro:	IPR003015:	HLH_Myc.		
DR	InterPro:	IPR001092:	HLH_dim.		
DR	InterPro:	IPR001321:	HypoxIndf1a.		
DR	InterPro:	IPR001610:	PAC.		
DR	InterPro:	IPR000014:	PAS.		
DR	Pfam:	PF00785:	PAC: 1.		
DR	Pfam:	PF00989:	PAS: 2.		
DR	PRINTS:	PRO1080:	HYPOXAIF1A.		
DR	SMART:	SMO0353:	HLH: 1.		
DR	SMART:	SMO0086:	PAC: 1.		
DR	SMART:	SMO0091:	PAS: 2.		
DR	PROSITE:	PS00038:	HELIX_LOOP_HELIX: 1.		
DR	PROSITE:	PS50112:	PAS: 2.		
DR	Repeat:	DNA-binding:	Nuclear protein; Transcription regulation;		
KW	Activator:	Phosphorylation.			
FT	DNA_BIND	17	30		
FT	DOMAIN	31	71		
FT	DOMAIN	80	155		
FT	DOMAIN	228	298		
FT	DOMAIN	302	345		
FT	CONFLICT	31	31		
FT	CONFLICT	128	128		
FT	CONFLICT	351	351		
FT	CONFLICT	511	511		
FT	CONFLICT	686	686		
FT	CONFLICT	785	785		
QO	SEQUENCE	822 AA:	91858 MW:	E717ADCECA9D795 CRC64;	

[illegible]

	RESULT	2
HIFA_HUMAN		
ID	HIFA_HUMAN	STANDARD; PRT; 826 AA.
AC	Q16665;	
DT	01-NOV-1997 (Rel. 35, Created)	
DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (ARNT interacting protein) (Member of PAS protein 1) (MOP1) (HIF1 alpha).	
GN	Hifa.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
OX	NCBI_TaxID:9606;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 166-170; 259-289 AND 771-781.	
RX	MEDLINE=95296340; PubMed=7539918;	
RA	Wang G.L., Jiang B.-H., Rue E.A., Semenza G.L.;	
RT	"Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 92:5510-5514(1995).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Hepatoma;	
RX	MEDLINE=87236817; PubMed=9079689;	
RA	Hogeneesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,	
RT	Przy-Grant M., Perdev G.H., Bradfield C.A.;	
RL	"Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway.";	
RL	J. Biol. Chem. 272:8581-8593(1997).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RA	Rupert J.L., Hochachka P.W.;	
RT	"Hifa sequence in the Quechua, a high altitude population.";	
RL	Submitted (NOV-1999) to the EMBL/genbank/DBJ databases.	
CC	-1- FUNCTION: INVOLVED IN THE INDUCTION OF OXYGEN REGULATED GENES SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE ELEMENT (HRE).	
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER BHLH PROTEIN. HETEROODIMER OF AN ALPHA AND A BETA (ARNT) SUBUNITS.	
CC	THE ALPHA SUBUNIT IS UNIQUE TO HIF-1 WHEREAS HIF-1 BETA (ARNT) CANNOT DIMERIZE WITH OTHER BHLH-PAS PROTEINS. INTERACTS WITH HSP90.	
CC	-1- SUBCELLULAR LOCATION: Nuclear (Potential).	
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES. HIGHEST LEVELS IN KIDNEY AND HEART.	
CC	-1- INDUCTION: UNDER REDUCED OXYGEN TENSION.	
CC	-1- DOMAIN: BOTH HYPOXIA RESPONSIVENESS AND TRANSACTIVATION CAPABILITY	



```

CC RESIDE WITHIN THE C-TERMINAL PART.
CC -1- PTV: REQUIRES PHOSPHORYLATION FOR DNA-BINDING.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U22431; AAC51052.1; -
DR EMBL: U29165; AAC51210.1; -
DR EMBL: AF207601; AAF20139.1; -
DR EMBL: AF207602; AAF20140.1; -
DR EMBL: AF208487; AAF20149.1; -
DR TRANSFAC: T01610; -
DR MIM: 603348; -
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR001321; Hypoxindf1A.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00989; PAS; 2.
DR PRINTS: PR01080; HYPOXIAIF1A.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX-LOOP-HELIX; 1.
DR PROSITE: PS00112; PAS; 2.
DR Repeat: DNA-binding; Nuclear protein; Transcription regulation;
DR Activator; Phosphorylation.
FT DNM_BIND 17 30 BASIC DOMAIN.
FT DOMAIN 31 71 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 85 158 PAS 1.
FT DOMAIN 228 298 PAS 2.
FT DOMAIN 302 345 PAC.
FT DOMAIN 615 621 POLY-THR.
SQ SEQUENCE 826 AA; 92670 MM; ABD4F7DAA135BE2D CRC64;

```

Query Match 30.38; Score 495; DB 1; Length 826;  
 Best Local Similarity 46.28; Pred. No. 2.3e-35;  
 Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;

```

OY 11 NTELRKESRAARSRSOETEVLYOLAHNLPPARGVSAHLDDKASIMLTSTYLMHRLC 70
DB 14 SSKRRKESRAARSRSKSESEVFEELAHQLPLPHNVSHLDKASVMLTSTYLMHRLC 73
OY 71 AAGKRGKATGRLLPEGGFGRHGRHGRGLPVGKCOQA-----PGPOSVLC 120
DB 74 DAG-----DLIEDMKMKOMCNFYLAKDGFVNLTD 105
OY 121 SSSLIH-----NPTPG-TNESLELGHSIFDFIHPDCELODALTPRPNISSKKLEAPT 174
DB 106 DGDIIYISDNVKNYGLQF--ELTGHVSFDFTHPCDHEEKREMLTHNGIYKKKEQONT 163
OY 175 ERHSFLMKSTLTSRGRTLNKAATMKVLHSGHMRABKPPAQSPPASPSSEPTLOCLV 234
DB 164 QRSFLLNKCTLTLSRGRTLNKAATMKVLHSGHMRABKPPAQSPPASPSSEPTLOCLV 221
OY 235 LICEAIP 241
DB 222 LICEPIP 228

```

RESULT 3  
 PAS1\_MOUSE

```

ID PAS1_MOUSE STANDARD: PRT; 874 AA.
AC P97481; 008787; 055046;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Endothelial PAS domain protein 1 (EPAS1), a transcription factor
DE (MHF) (HIF-related factor) (HRF).
GN EPAS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RX MEDLINE=97152468; PubMed=9000051;
RA Tian H., McKnight S.L., Russell D.W.;
RT "Endothelial PAS domain protein 1 (EPAS1), a transcription factor
RT selectively expressed in endothelial cells.";
RL Genes Dev. 11:72-82(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Hypothalamus, and Skeletal muscle;
RX MEDLINE=97272213; PubMed=9113979;
RA Ema M., Taya S., Yokotani N., Sogawa K., Matsuda Y.,
RA Fujii-Kuriyama Y.;
RT "A novel BHLH-PAS factor with close sequence similarity to hypoxia-
RT inducible factor 1alpha regulates the VEGF expression and is
RT potentially involved in lung and vascular development.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:4273-4278(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain capillary;
RX MEDLINE=97321546; PubMed=9178256;
RA Flamme I., Froehlich T., von Reutern M., Kappel A., Damerat A.,
RA Risaun W.;
RT "HRF, a putative basic helix-loop-helix-PAS-domain transcription
RT factor is closely related to hypoxia-inducible factor-1 alpha and
RT developmentally expressed in blood vessels.";
RL Mech. Dev. 63:51-60(1997).
CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE INDUCTION OF OXYGEN
CC REGULATED GENES. SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE
CC ELEMENT (HRE). REGULATES THE VASCULAR ENDOTHELIAL GROWTH FACTOR
CC (VEGF) EXPRESSION AND SEEMS TO BE IMPLICATED IN THE DEVELOPMENT OF
CC BLOOD VESSELS AND THE TUBULAR SYSTEM OF LUNG. MAY ALSO PLAY A ROLE
CC IN THE FORMATION OF THE ENDOTHELIUM GIVING RISE TO THE BLOOD BRAIN
CC BARRIER. POTENT ACTIVATOR OF THE TIE-2 TYROSINE KINASE EXPRESSION.
CC BHLH PROTEIN. HETERODIMER WITH THE ARNT PROTEIN.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. HETERODIMER WITH THE ARNT PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, WITH HIGHEST LEVELS
CC IN LUNG, FOLLOWED BY HEART, KIDNEY, BRAIN AND LIVER. PREDOMINANTLY
CC EXPRESSED IN ENDOTHELIAL CELLS. ALSO FOUND IN SMOOTH MUSCLE CELLS
CC OF THE UTERUS, NEURONS, AND BROWN ADIPOSE TISSUE. HIGH EXPRESSION
CC IN EMBRYONIC CHOROID PLEXUS AND KIDNEY GLOMERULI.
CC -1- DEVELOPMENTAL STAGE: IN DAY 11 EMBRYO, EXPRESSION IS ALMOST
CC EXCLUSIVELY SEEN IN ENDOTHELIAL CELLS OF THE INTERSEGMENTAL BLOOD
CC VESSELS SEPARATING THE SOMITES, THE ATRIAL AND VENTRICULAR
CC CHAMBERS OF THE HEART, AND THE DORSAL AORTA. HIGH EXPRESSION ALSO
CC OCCURS IN EXTRA-EMBRYONIC MEMBRANES. IN THE DEVELOPING BRAIN OF
CC DAY 13 EMBRYO, ENDOTHELIAL CELLS OF THE HIGHLY VASCULARIZED
CC CHOROID PLEXUS CONTAIN HIGH LEVELS OF EPAS1.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



FT	DOMAIN	84	154	PAS 1.
FT	DOMAIN <td>230</td> <td>300</td> <td>PAS 2.</td>	230	300	PAS 2.
FT	DOMAIN <td>304</td> <td>347</td> <td>PAC.</td>	304	347	PAC.
FT	DOMAIN <td>474</td> <td>480</td> <td>POLY-SER.</td>	474	480	POLY-SER.
FT	CONFLICT <td>60</td> <td>60</td> <td>A -&gt; E (IN REF. 1).</td>	60	60	A -> E (IN REF. 1).
FT	CONFLICT <td>539</td> <td>539</td> <td>G -&gt; D (IN REF. 1).</td>	539	539	G -> D (IN REF. 1).
FT	CONFLICT <td>601</td> <td>601</td> <td>R -&gt; H (IN REF. 1).</td>	601	601	R -> H (IN REF. 1).
FT	CONFLICT <td>693</td> <td>693</td> <td>N -&gt; D (IN REF. 1).</td>	693	693	N -> D (IN REF. 1).
FT	CONFLICT <td>716</td> <td>716</td> <td>K -&gt; E (IN REF. 1).</td>	716	716	K -> E (IN REF. 1).
FT	CONFLICT <td>722</td> <td>722</td> <td>P -&gt; L (IN REF. 1).</td>	722	722	P -> L (IN REF. 1).
FT	CONFLICT <td>765</td> <td>765</td> <td>L -&gt; F (IN REF. 1).</td>	765	765	L -> F (IN REF. 1).
FT	CONFLICT <td>769</td> <td>769</td> <td>S -&gt; P (IN REF. 1).</td>	769	769	S -> P (IN REF. 1).
FT	CONFLICT <td>844</td> <td>844</td> <td>R -&gt; C (IN REF. 1).</td>	844	844	R -> C (IN REF. 1).
FT	CONFLICT <td>847</td> <td>847</td> <td>K -&gt; N (IN REF. 1).</td>	847	847	K -> N (IN REF. 1).
SO	SEQUENCE <td>870 AA;</td> <td>96425 MW;</td> <td>30F5B7B13AEC871D CRC64;</td>	870 AA;	96425 MW;	30F5B7B13AEC871D CRC64;
Query Match				
Best Local Similarity		28.5%;	Score 467;	DB 1; Length 870;
Matches 111;		Conservative 43.9%;	Pred. No. 6.7e-33;	
			Mismatches 82;	Indels 26; Gaps 8;
OY	1	MALGIQVRNSTELTKKESRDAARRSQETFEVLYQLAHTLPFANGVSAHLDKASIMRLT	60	
Db	1	MTADEKKRRSSERRKESRDAARRSRKSETEVFYELAHELPLPHSVSHLDKASIMRLA	60	
OY	61	ISYLEKHL---CAAGKGRGATGRLLPEGGGGRHGHGRGRGLPGKQQAQPPQS	116	
Db	61	ISFLTKHLSSVGCENSESEAEADQDQ-----DNLTKALEGF-----IAVTD	105	
OY	117	VDLCSSSLHNPTGTFNFSLELIGHISIFDFIHPDQDEIADALTPR--PNLSKKRLAPT	174	
Db	106	GDMIFLSNISIKFMGLT-QVELTGHISIFDFIHPDQDEIETRENLSTKNGSGFGKSKDMST	164	
OY	175	ERHEFLRKSTLTSGRLNKAATWXYLHCGSHMRAYKP-PAQNSPAGSPSPSEPPLOLT	233	
Db	165	EDDFMRKCVYTNNGRTVNLKSAITWYLDHCTGVKNVNNCPHNSLGG--YKEPLLSTL	222	
OY	234	VLICEAIPQLPDPH	246	
Db	223	ITMCEPI-QHPESH	234	
RESULT 5				
NPAL_HUMAN	STANDARD;	PRT;	590 AA.	
ID	NPAL_HUMAN			
AC	O99742; Q99632;			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Neuronal PAS domain protein 1 (Neuronal PAS1) (Member of PAS protein 5) (MOP5).			
GN	NPAS1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=97165088; PubMed=9012850;			
RA	Zhou Y.-D., Barnard M., Tian H., Li X., Ring H.Z., Francke U.,			
RA	Shelton J., Richardson J., Russell D.W., McNight S.L.,			
RT	"Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 94:713-718(1997).			
RN	[2]			
RP	SEQUENCE OF 110-590 FROM N.A.			
RC	TISSUE=Hepatoa;			
RX	MEDLINE=97236817; PubMed=9079689;			
RA	Hogensch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,			
RA	Pray-Grant J.B., Perdew G.H., Bradfield C.A.;			
RT	"Characterization of a subset of the basic-helix-loop-helix-PAS			
RT	superfamily that interacts with components of the dioxin signaling			

[illegible]

```
RESULT 6
NPA1_MOUSE STANDARD; PRT; 594 AA.
ID NPA1_MOUSE
AC P97459;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuronal PAS domain protein 1 (Neuronal PAS1).
GN NPA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97165088; PubMed=9012850;
RA Zhou Y.-D., Barnard M., Tian H., Li X., Ring H.Z., Francke U.,
RA Shelton J., Richardson J., Russell D.W., McKnight S.L.;
RT "Molecular characterization of two mammalian bHLH-PAS domain proteins
RT selectively expressed in the central nervous system.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:713-718(1997).
CC -1 SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN.
CC -1 SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1 TISSUE SPECIFICITY: IN BRAIN, EXCLUSIVELY NEURONAL. ALSO FOUND IN
CC SPINAL CORD.
CC -1 DEVELOPMENTAL STAGE: FIRST DETECTED BETWEEN EMBRYONIC DAY 15 AND
CC DAY 16.
CC -1 SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -1 SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1 SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR MGD; U77967; AAB47247.1; -.
DR EMBL; MGI:109205; Npas1.
DR InterPro; IPR003015; HLH_Myc.
DR InterPro; IPR001092; HLH_dim.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS.
DR Pfam; PF00989; PAS; 1.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS00036; HELIX_LOOP_HELIX; 1.
DR PROSITE; PS00112; PAS; 1.
DR Repeat; DNA-binding; Nuclear protein; Transcription regulation.
KW DNA BIND.
FT DNA BIND 46 58 BASIC DOMAIN.
FT DOMAIN 59 99 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 135 205 PAS 1.
FT DOMAIN 294 360 PAS 2.
FT DOMAIN 366 409 PAC.
FT DOMAIN 212 223 POLY-SER.
SQ SEQUENCE 594 AA; 63736 MW; D6477696DF69A4B3 CRC64;
```

```
Query Match 17.9%; Score 293.5; DB 1; Length 594;
Best Local Similarity 33.1%; Pred. NO. 5.1e-18;
Matches 101; Conservative 32; Mismatches 63; Indels 109; Gaps 15;
```

```
OY 15 RKESSRAARSRSGEFTVLQALHTLPFARGVSADLKASIMRLTISYLMHNRCAAGG 74
DB 46 RKESSRAARSRSGEFTVLQALHTLPFARGVSADLKASIMRLTISYLMHNRCAAGG 105
```

```
OY 75 KRQ--RATGRLLPEGCGFRHGRHGRHGLPYGKQQAQP---QSYDLCSLLHNPT 129
DB 106 PPMGLRAVG-----PPAGLAPG--RRGPVALVEVEFOHGLITLQSLD----- 147
OY 130 RGTNFSL-----ELIGHSIFDFTHPCDQBELQDAL----- 159
DB 148 -GRVFALNDGKFLYISETVSYTLGLSQVELTSSVFDDITHPDHSEVLEQLGLRAASIG 206
OY 160 --TPRPNLRSKK-----KLEA-PT-----ERHPSLRKSKTSLTSGRFTLNT 195
DB 207 PPTP-PSVSSSSSSSSSSSLVDPTPEIASPTASPARFAQGSFVNRKSTLTKRG--LNV 263
OY 196 KAATWKVLCGSHMRVK-----PPAQTSPPAGSPSEPPLO-----CLV 234
DB 264 KASGKYVHWYGRRLRARALGLVALGHTLPPA-----PLAELPLGHMIVFRLSLGLTI 316
OY 235 LICEA 239
DB 317 LACES 321
RESULT 7
STRA-DROME STANDARD; PRT; 1507 AA.
ID STRA-DROME
AC Q24167; Q9VAA5;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Similar protein.
GN STRA OR CG7951.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96269413; PubMed=8682312;
RA Nambu J.R., Chen W., Hu S., Crews S.T.;
RT "The Drosophila melanogaster similar bHLH-PAS gene encodes a protein
RT related to human hypoxia-inducible factor 1 alpha and Drosophila
RT single-minded.";
RL Gene 172:249-254 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li F.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Andrews-Pfankoch C., Baldwin D.,
RA Abail J.F., Abghayani A., An H.-J., Bayraktaroglu L., Beasley E.M.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kratiz C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
```

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sviderskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Welschbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Yeh J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: POSSIBLE DNA-BINDING TRANSCRIPTIONAL ACTIVATOR.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN THE EMBRYO.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: UA3090; AAC47303.1; -.  
 DR EMBL: AE003772; AAF57008.2; -.  
 DR FlyBase: FBgn0015542; slma.  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR000014; PAC.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE: PS50112; PAS; 2.  
 KW Repeat; DNA-binding; Nuclear protein; Transcription regulation;  
 KW Activator; Coiled coil.  
 KW DNA\_BIND 72 85  
 FT DOMAIN 126 126 BASIC DOMAIN.  
 FT DOMAIN 167 240 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 307 377 PAS 1.  
 FT DOMAIN 381 422 PAS 2.  
 FT DOMAIN 577 587 PRO-RICH.  
 FT DOMAIN 880 908 COILED COIL (POTENTIAL).  
 FT DOMAIN 982 1054 COILED COIL (POTENTIAL).  
 FT DOMAIN 1110 1162 COILED COIL (POTENTIAL).  
 FT DOMAIN 26 39 POLY-SER.  
 FT DOMAIN 718 725 POLY-SER.  
 FT DOMAIN 759 763 POLY-SER.  
 FT DOMAIN 767 776 POLY-SER.  
 FT DOMAIN 907 918 POLY-SER.  
 FT DOMAIN 945 948 POLY-SER.  
 FT DOMAIN 990 998 POLY-SER.  
 FT DOMAIN 1020 1038 POLY-SER.  
 FT DOMAIN 1113 1126 POLY-SER.  
 FT DOMAIN 1146 1162 POLY-SER.  
 FT DOMAIN 1205 1208 POLY-SER.  
 FT DOMAIN 1277 1284 POLY-SER.  
 FT DOMAIN 1298 1301 POLY-SER.  
 FT CONFLICT 38 38 S -> A (IN REF. 1).  
 FT CONFLICT 345 345 S -> L (IN REF. 1).  
 FT CONFLICT 492 492 A -> V (IN REF. 1).  
 FT CONFLICT 588 588 T -> I (IN REF. 1).

FT CONFLICT 709 709 T -> R (IN REF. 1).  
 FT CONFLICT 776 776 Q -> Q (IN REF. 1).  
 FT CONFLICT 895 895 Q -> Q (IN REF. 1).  
 FT CONFLICT 902 902 Q -> S (IN REF. 1).  
 FT CONFLICT 982 982 A -> T (IN REF. 1).  
 FT CONFLICT 1125 1126 MISSING (IN REF. 1).  
 FT CONFLICT 1154 1157 MISSING (IN REF. 1).  
 FT CONFLICT 1444 1444 F -> L (IN REF. 1).  
 FT CONFLICT 1447 1447 G -> C (IN REF. 1).  
 FT CONFLICT 1451 1451 S -> N (IN REF. 1).  
 FT CONFLICT 1494 1494 D -> G (IN REF. 1).  
 SQ SEQUENCE 1507 AA; 165822 MW; 4102939C8FBF006 CRC64;  
 Query Match 17.2%; Score 281; DB 1; Length 1507;  
 Best Local Similarity 30.2%; Pred. No. 1.9e-16;  
 Matches 71; Conservative 47; Mismatches 71; Indels 46; Gaps 5;  
 QY 9 RSNTELRKRSRDARSRSETEVLYQALHTPFGANGSAHLDPKASIMRLTSLMHR 68  
 DB 67 RNNKRRKRSRDARSRSETEVLYQALHTPFGANGSAHLDPKASIMRLTSLMHR 126  
 QY 69 L-----CAAGKRGKRGATGRLLPDGPGRHGRHGRHGLPVKCOQAPGQSVLD 119  
 DB 127 MLDQVPSLRDNDIKQDIETAE-----DQGEVVKLEVTGEMDLNCAEAREL 174  
 QY 120 CSSSL-----INNPRTGTFNS-----LELGHSITFDIHPDDELDALTPR 162  
 DB 175 LKQTMDFLVLVSHEDGITYSENVVEYLGITKIDTIGQOITWESHOCDNAIEALSLK 234  
 QY 163 PNLSSKRLKAPTE-----RHSFLMKSTLTSRGRPLNKAATWVLCQSGHM 209  
 DB 235 RELAQKVKDEPQNSGVSTHHRDLFVRLKCTLISRGKSINIKSAYIVHITGHL 289  
 RESULT 8  
 SIML\_MOUSE STANDARD; PRT; 765 AA.  
 ID SIML\_MOUSE  
 AC 061045; P70183;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Single-minded homolog 1 (MSIM1).  
 GN SIM1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS WEBSTER;  
 RX MEDLINE=97020303; PubMed=8812055;  
 RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,  
 RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,  
 RA Tessier-Lavigne M.;  
 RT "Expression patterns of two murine homologs of *Drosophila*  
 RT single-minded suggest possible roles in embryonic patterning and in  
 RT the pathogenesis of Down syndrome.";  
 RT Mol. Cell. Neurosci. 7:1-16(1996).  
 RN [2]  
 RP ERRATUM.  
 RX MEDLINE=97029422; PubMed=8875433;  
 RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,  
 RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,  
 RA Tessier-Lavigne M.;  
 RL Mol. Cell. Neurosci. 7:519-519(1996).  
 RN [3]  
 RP REVISIONS TO C-TERMINUS.  
 RX MEDLINE=97343329; PubMed=9199934;  
 RA Fan C.-M.;  
 RL Unpublished results, cited by:  
 RL Christ R., Scott H.S., Chen H., Kudoh J., Rossier C., Minoshima S.,  
 RL Wang Y., Shimizu N., Antonarakis S.E.;

RL Genome Res. 7:615-624(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6J;  
 RX MEDLINE-96413339; PubMed-8927054;  
 RA Ema M., Morita M., Ikawa S., Tanaka M., Matsuda Y., Gotoh O.,  
 RA Saijoh Y., Fujii H., Hamada H., Kikuchi Y., Fujii-Kuriyama Y.;  
 RT "Two new members of the murine Slm gene family are transcriptional  
 RT repressors and show different expression patterns during mouse  
 RT embryogenesis";  
 RL Mol. Cell. Biol. 16:5865-5875(1996).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-129/SV;  
 RX Hosoya T.;  
 RN Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.  
 RP SUBUNIT.  
 RX MEDLINE-97172525; PubMed-9020169;  
 RA Probst M.R., Fan C.M., Tessier-Lavigne M., Hankinson O.;  
 RT "Two murine homologs of the Drosophila single-minded protein that  
 RT interact with the mouse aryl hydrocarbon receptor nuclear  
 RT translocator protein";  
 RL J. Biol. Chem. 272:4451-4457(1997).  
 CC - FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS  
 CC DURING EMBRYOGENESIS AND IN THE ADULT.  
 CC - SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER OF SIM1 AND ARNT.  
 CC - TISSUE SPECIFICITY: DETECTED IN LUNG, SKELETAL MUSCLE AND KIDNEY.  
 CC DURING FETAL DEVELOPMENT IT IS FOUND IN THE CNS, DEVELOPING  
 CC KIDNEY, MESODERMAL AND ENDODERMAL TISSUES, INCLUDING DEVELOPING  
 CC SOMITES, MESONEPHRIC DUCT, AND FOREGUT.  
 CC - SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC - SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC - SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U40575; AAA91201.1; ALT\_SEQ.  
 DR EMBL: D79209; BAA11467.1; -.  
 DR EMBL: AB013491; BAA28270.1; -.  
 DR EMBL: AB013484; BAA28270.1; JOINED.  
 DR EMBL: AB013485; BAA28270.1; JOINED.  
 DR EMBL: AB013486; BAA28270.1; JOINED.  
 DR EMBL: AB013487; BAA28270.1; JOINED.  
 DR EMBL: AB013488; BAA28270.1; JOINED.  
 DR EMBL: AB013489; BAA28270.1; JOINED.  
 DR EMBL: AB013490; BAA28270.1; JOINED.  
 DR MGI: 96306; Sim1.  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001092; HLH\_dlm.  
 DR InterPro: IPR001067; Nucleinslocator.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR PRINTS: SM00785; NCTRNSLOCATR.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE: PS50112; PAS; 2.  
 DR Developmental protein; Neurogenesis; Nuclear protein; Repeat;  
 KW Transcription regulation; DNA-binding.  
 FT DNA\_BIND 1 13 BASIC DOMAIN  
 FT DOMAIN 14 54 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).

FT DOMAIN 77 147 PAS 1.  
 FT DOMAIN 218 288 PAS 2.  
 FT CONFLICT 133 133 H -> L (IN REF. 1).  
 FT CONFLICT 176 176 MISSING (IN REF. 1).  
 FT CONFLICT 322 322 P -> R (IN REF. 1).  
 FT CONFLICT 480 480 A -> P (IN REF. 1).  
 FT CONFLICT 537 537 D -> S (IN REF. 1).  
 SQ SEQUENCE 765 AA; 85540 MW; B1AF7DA8578CD17 CRC64;  
  
 Query Match 16.4%; Score 268; DB 1; Length 765;  
 Best Local Similarity 36.2%; Pred. No. 1,1e-15;  
 Matches 77; Conservative 28; Mismatches 86; Indels 22; Gaps 6;  
  
 QY 16 KEKSPAAARSRSQETEVYQLAHLTPFAGVSAHLDKASIMRLTJSTYLMHRICAAAG-G 74  
 DB 2 KEKSKNAARTRREKENSEFEELAKLPLPSAITSQDKASTILNTSYLMKRVPPEDGIG 61  
 QY 75 KRGKATGRLPEGPGRHGTN-RRGRHGL-----PVKCOAPGPOSDVLCSSSLIHPN 128  
 DB 62 EAMGHTRSRSPLDNVGRELSHLQTLDFIFVAVPDGKIMYISRTASVHLGSL----- 115  
 QY 129 TPQTNLSLELGHSIFDFHPCQDELQDALTP-RNLSKKLEAFTERHFSLRMSTLT 187  
 DB 116 -----QVELTGNISYIEYHPADHDMTAVLTAAHQPHSHFVQEYIEERFSFLRMKVLA 169  
 QY 188 SRGRTLNKAATKWLKVGSGHMRAPKPPAQTSP 220  
 DB 170 K--RNAGLICGKIVHCISGLKIRQYSLDMSP 200  
  
 RESULT 9  
 SIM1 HUMAN STANDARD; PRT; 766 AA.  
 ID SIM1 HUMAN  
 AC P81133;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Single-minded homolog 1.  
 GN SIM1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-97343329; PubMed-9199934;  
 RA Christ R., Scott H.S., Chen H., Kudoh J., Rossler C.,  
 RA Minoshima S., Wang Y., Shimizu N., Antonarakis S.E.;  
 RT "Cloning of two human homologs of the Drosophila single-minded gene  
 RT SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome  
 RT chromosomal region";  
 RL Genome Res. 7:615-624(1997).  
 CC - FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS  
 CC DURING EMBRYOGENESIS AND IN THE ADULT.  
 CC - SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER OF SIM1 AND ARNT.  
 CC - SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC - SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC - SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U70212; AAB62395.1; -.  
 DR MIM: 603128; -.  
 DR InterPro: IPR003015; HLH\_Myc.







[illegible]

Query Match	15.68;	Score	256;	DB	1;	Length	667;
Best Local Similarity	35.38;	Pred. No.	1.1e-14;				
Matches	73;	Conservative	26;	Mismatches	78;	Indels	30;
						Gaps	6

```
QY      16 KEKRDAAARSRSOETEVVLQVLAATLPFRAGVSANHIDKASIMULTSYLRMHNLCAGGK   75
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      2 KKKNNNAKTTRREKEENGEFFELAKLTPLSAITSSODDKASIIILTLTYSYLKMRAWPEEG--   59

QY      76 RGRATGRLLPREGPGCFRHHGTNRRCNGLPV-----GKCQDAAPGPOSDLCSSU    124
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      60 LGDAMGQ--PSRAGRDLGVAKELGSNLLQLTLDGFVFNVAADGKIIMTISELAVNHLS--    115

QY     125 IHNPTGTFNSFLDGHSIFDEIHPDCOELODALTPRWISKKKL-EAETRNFSILRK    183
        ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db     116 -----QVELTGMSIYEIYIPRSHDEFTAVLNQRLLNHLNLQEYETIESFGLRYK    165

QY     184 STTSRGFTLMKAATKYLVLHCSGHMIR 210

Db     166 CVLAR--RNAAGTCGGYKRVIHCSGYLL 190
```

RESULT	12
TRH_DROME	
ID	TRH_DROME
AC	Q24119; Q24165; Q9W0Q7;
DT	01-NOV-1997 (Rel. 35, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Trachealeless protein.
GN	TRH OR CG6883.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxId=7227;
RM	[1]
RP	SEQUENCE FROM N.A.
RP	TISSUE=Embryo;
RC	MEDLINE=96136712; PubMed=8557198;
RX	Wilk R., Weizman I., Shilo B. -2.;
RA	"Trachealeless encodes a bHLH-PAS protein that is an inducer of
RT	tracheal cell fates in Drosophila.";
RL	Genes Dev. 10:93-102(1996).

[2]  
SEQUENCE FROM N.A.  
RP TISSUE=Embryo;  
RC MEDLINE=96136713; PubMed=8557189;  
RX Isaac D.D., Andrew D.J.;  
RT "Tubulogenesis in Drosophila: a requirement for the trachealless gene product.", 10:103-117(1996).  
RL Genes Dev. 10:103-117(1996).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC STRAIN=BERKELEY.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Rge Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George A.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Burton R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mkios G.L.G.,  
RA Abdl J.F., Agbayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
RA Bokkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dublin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegwan C.,  
RA Jostal M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Schelder F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
CC -1- FUNCTION: TRANSCRIPTION FACTOR. MASTER REGULATOR OF TRACHEAL CELL  
CC FATES IN THE EMBRIO, NECESSARY FOR THE DEVELOPMENT OF THE SALIVARY  
CC GLAND DUCT AND THE POSTERIOR SPIRACLES. IT MAY INDUCE A GENERAL  
CC FATE OF BRANCHED TUBULAR STRUCTURES OF EPITHELIAL ORIGIN. TGO/TRH  
CC HETERODIMERS ARE INVOLVED IN THE CONTROL OF BREATHLESS EXPRESSION  
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
CC BHLH PROTEIN, HETERODIMER WITH TGO.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2 AND  
CC 3; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: TRACHEAL, SALIVARY GLAND DUCTS, POSTERIOR  
CC SPIRACLES (FILIKOPER PRIMORDIA) AND A SUBSET OF CELLS IN THE CNS.  
CC -1- DEVELOPMENTAL STAGE: DURING EMBRYOGENESIS, FIRST DETECTED IN THE  
CC TRACHEAL PLACODES AT STAGE 8, AND EXPRESSION CONTINUES THROUGHOUT  
CC EMBRYONIC AND LARVAL DEVELOPMENT. IN THE DEVELOPING SALIVARY  
CC GLAND, EXPRESSION IS OBSERVED IN THE ENTIRE GLAND AT STAGE 9 AND  
CC BY STAGE 12. EXPRESSION IS CONFINED TO THE SALIVARY DUCTS.  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS.  
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
CC -1- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS  
CC IN POSITIONS 22 TO 34.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U33427; AAA96257.1; ALT\_INIT.  
DR EMBL: U42699; AAA96754.1; ALT\_FRAME.  
DR EMBL: AE003468; AAR47386.1; -.  
DR EMBL: AE003468; AAR47386.1; trh.  
DR InterPro: IPR003015; HLH\_MYC.  
DR InterPro: IPR001092; HLH\_dim.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.  
DR Pfam: PF00785; PAC; 1.  
DR Pfam: PF00989; PAS; 2.  
DR SMART: SM00353; HLH; 1.  
DR SMART: SM00086; PAC; 1.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
DR PROSITE: PS0112; PAS; 2.  
KW Developmental protein; Nuclear protein; Transcription regulation;  
KW Repeat; DNA-binding; Alternative splicing.  
FT DNA\_BIND 86 99 BASIC DOMAIN.  
FT DOMAIN 100 140 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT DOMAIN 174 244 PAS 1.  
FT DOMAIN 391 461 PAS 2.  
FT DOMAIN 465 508 PAC.  
FT DOMAIN 629 636 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 154 157 POLY-SER.  
FT DOMAIN 244 250 POLY-GLY.  
FT DOMAIN 251 254 POLY-SER.  
FT DOMAIN 255 264 POLY-GLN.  
FT DOMAIN 722 728 POLY-SER.  
FT VARSPLIC 281 286 MISSING (IN ISOFORM 2).  
FT VARSPLIC 328 356 MISSING (IN ISOFORM 3).  
FT CONFLICT 78 78 G -> A (IN REF. 1).  
FT CONFLICT 250 250 P -> G (IN REF. 2).  
FT CONFLICT 703 703 A -> T (IN REF. 1).  
FT CONFLICT 708 708 A -> P (IN REF. 1).  
FT CONFLICT 829 829 A -> V (IN REF. 1).  
SQ SEQUENCE 958 AA; 102241 MW; 8F9CF58E1370541 CRC64;  
Query Match 15.3%; Score 251; DB 1; Length 958;  
Best Local Similarity 27.2%; Pred. No. 4.5e-14;  
Matches 89; Conservative 41; Mismatches 91; Indels 106; Gaps 10;

Db 371 SHSRKSPPLGMAVALATLPPPSVHE 397  
RESULT 13  
SIM\_DROME STANDARD; PRT; 673 AA.  
AC P05709; Q9VPE23;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Single-minded protein.  
GN SIM OR CG7771.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE OF 19-673 FROM N.A.  
RX MEDLINE=88151023; PubMed=3345560;  
RA Crews S.T., Thomas J.B., Goodman C.S.;  
RT "The Drosophila single-minded gene encodes a nuclear protein with  
RT sequence similarity to the per gene product.";  
RL Cell 52:143-151(1988).  
RN [2]  
RP SEQUENCE OF 1-18 FROM N.A., AND SIMILARITY TO HLH PROTEINS.  
RX MEDLINE=92103681; PubMed=1760843;  
RA Nambu J.R., Lewis J.O., Wharton K.A. Jr., Crews S.T.;  
RT "The Drosophila single-minded gene encodes a helix-loop-helix protein  
RT that acts as a master regulator of CNS midline development.";  
RL Cell 67:1157-1167(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99054545; PubMed=9840810;  
RA Kasai Y., Stahl S., Crews S.;  
RT "Specification of the Drosophila CNS midline cell lineage: direct  
RT control of single-minded transcription by dorsal/ventral patterning  
RT genes.";  
RL gene Expr. 7:171-189(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,  
RA Butkus K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Duderin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon S., Pollard J., Puri V., Reese M.G.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massarman D.A., Weinstock G.M., Weissensbach J.,  
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*,"  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT FUNCTIONS AS A MASTER  
 CC DEVELOPMENTAL REGULATOR OF THE CNS MIDLINE LINEAGE. MUTATIONS IN  
 CC THE SIM GENE RESULTS IN THE LOSS OF THE PRECURSOR CELLS GIVING  
 CC RISE TO MIDLINE CELLS OF THE EMBRYONIC CENTRAL NERVOUS SYSTEM.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M19020; AAA28900.1; -  
 DR EMBL: AF071934; AAC64519.1; ALT\_SEQ.  
 DR EMBL: AE003698; AAF54902.1; ALT\_SEQ.  
 DR PIR: A29945; A29945.  
 DR TRANSFAC: T00750; -  
 DR FLYBase: FBgn000466; sim.  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001092; HLH\_dlm.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM0086; PAC; 1.  
 DR SMART: SM0086; PAC; 2.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE: PS0112; PAS; 2.  
 KW Developmental protein; Neurogenesis; Nuclear protein; Repeat;  
 KW Transcription regulation; DNA-binding.  
 FT DNAS\_BIND 1 13  
 FT DOMAIN 14 54 BASIC DOMAIN.  
 FT DOMAIN 76 148 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 242 312 PAS 1.  
 FT DOMAIN 382 422 PAS 2.  
 FT DOMAIN 423 426 14 X 3 AA REPEATS OF A-A-Q.  
 FT DOMAIN 483 486 POLY-VAL.  
 FT DOMAIN 489 488 POLY-SER.  
 FT DOMAIN 492 492 POLY-ASN.  
 FT DOMAIN 496 499 POLY-HIS.  
 FT DOMAIN 501 505 POLY-GLN.  
 FT DOMAIN 529 533 POLY-SER.  
 FT DOMAIN 573 576 POLY-ASN.  
 FT DOMAIN 604 607 POLY-SER.  
 FT DOMAIN 649 669 GLN/HIS-RICH.  
 FT DOMAIN 127 127 I -> Y (IN REF. 3).  
 FT CONFLICT 401 409 MISSING (IN REF. 4).  
 SO SEQUENCE 673 AA; 73589 MW; 2F9FOABBA2BC0FBE CRC64;

Query Match 15.2% Score 249; DB 1; Length 673;  
 Best Local Similarity 30.1% Pred. No. 4.4e-14;  
 Matches 83; Conservative 39; Mismatches 84; Indels 70; Gaps 10;

OY 16 KEKRDARSRSQETVLYQLAHTLPARAGVSAHLDKASIMRLTISTLRHRLCAAGK 75  
 DB 2 KEKSNNAIRREKNEFCELAATLPAPAITOLDKASVIRLTSTYLRKROVPPDG-- 59

OY 76 RGRATGRLLPEGPCGFRHGRHRR-GRHGL-----PVKCGQAPGPOVDLCSSS 123  
 DB 60 LGFANG-----SSPAAQGRATIKELGSHLQLDGRIFVYVADGKIMYSERASVHLGLS- 114  
 OY 124 LIHNPTGNTSFLNLSIGHSIFDFHPCQDEELQDALTPRPMLSKRL----- 170  
 DB 115 -----QVELGNSIFEYIHNYDQDEMNALISLPHINQPLAOTHTPIGSPNGV 163  
 OY 171 EAPF-----ERHSLRKSKTLTSGRLNLNKAATKYLHSCGHHRAKPPAO 217  
 DB 164 QHPSAYDHDGRSHTEIKTEFFLRKMCVLAR--RAAGLTTSGFVKVHSGYLAIRIYDR 221  
 OY 218 TSPAGSPSEPPLQCLVLC-----EAPLPDPH 246  
 DB 222 GDGGS-----LIQNLGLVANGHSILPSSAITFEIKH 252  
 RESULT 14  
 CYCL\_DROME STANDARD: PRT: 413 AA.  
 AC 061734; 076344; 09YV44;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cycle protein (Brain and muscle ARNT-like 1) (BMAL1) (MOP3).  
 GN CYC OR CG8727.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Ovary;  
 RX MEDLINE=98292178; PubMed=9630224;  
 RA Rutisha J.E., Suri V., Le M., So W.V., Rosbash M., Hall J.C.;  
 RT "CYCLE is a second bHLH-PAS clock protein essential for circadian  
 RL rhythmically and transcription of *Drosophila* period and timeless.";  
 RL Cell 93:805-814(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Brain, and Muscle;  
 RX MEDLINE=98279147; PubMed=9616122;  
 RA Darlington T.K., Mager-Smith K., Ceriani M.F., Stekakis D., Gekakis N.,  
 RT Steeves T.D.L., Weitz C.J., Takahashi J.S., Kay S.A.;  
 RT "Closing the circadian loop: CLOCK-induced transcription of its own  
 RT inhibitors per and tim.";  
 RT Science 280:1599-1603(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abri J.F., Abhyani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dev I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,  
 RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,





...

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2002, 07:34:02 ; Search time 62.38 Seconds  
(without alignments)  
851.386 Million cell updates/sec

Title: US-09-896-791b-3  
Perfect score: 1636  
Sequence: 1 MALCIGRVSRNTLEKRSK.....TESSLPSVWLNRRKNCPC 307

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp.\_archaea:\*  
2: sp.\_bacteria:\*  
3: sp.\_fungi:\*  
4: sp.\_human:\*  
5: sp.\_invertebrate:\*  
6: sp.\_mammal:\*  
7: sp.\_mhc:\*  
8: sp.\_organelle:\*  
9: sp.\_phage:\*  
10: sp.\_plant:\*  
11: sp.\_rodent:\*  
12: sp.\_virus:\*  
13: sp.\_vertebrate:\*  
14: sp.\_unclassified:\*  
15: sp.\_virus:\*  
16: sp.\_bacteriap:\*  
17: sp.\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	855.5	52.3	630	11 Q9QX54	Q9qX54 mus musculu
2	855.5	52.3	662	11 Q9Z215	Q9z215 mus musculu
3	831	50.8	632	4 Q96K34	Q96K34 homo sapien
4	820.5	50.2	662	11 Q9JHS2	Q9jhs2 rattus norv
5	802	49.0	667	4 Q9Y2N7	Q9y2n7 homo sapien
6	573	35.0	648	4 Q9HAI2	Q9hai2 homo sapien
7	518.5	31.7	199	4 Q9HAM3	Q9ham3 mus saplen
8	503	30.7	811	13 Q9Y1B9	Q9y1b9 gallus gall
9	499	30.5	823	6 Q9XTAS	Q9xtas bos taurus
10	497	30.4	825	11 Q35800	Q35800 rattus norv
11	495	30.3	828	11 Q9CYA8	Q9cyas mus musculu
12	495	30.3	735	4 Q96PT9	Q96pt9 homo sapien
13	495	30.3	826	4 Q9UPB1	Q9upb1 homo sapien
14	491	30.0	823	11 Q9WT09	Q9wts1 rattus norv
15	482	29.5	874	11 Q9JHS1	Q9jhs1 rattus norv
16	471	28.8	867	13 Q9W7C6	Q9w7c6 gallus gall

17	471	28.8	870	13 Q9PTB3	Q9ptb3 coturnix co
18	468	28.6	870	6 Q9XTA4	Q9xta4 bos taurus
19	450	27.5	766	13 Q9BSW2	Q9bsw2 oncorhynch
20	426	26.0	237	4 Q9H7Z9	Q9h7z9 homo sapien
21	322	19.7	115	4 Q9S262	Q9s262 homo sapien
22	312	19.1	112	4 Q9UPH7	Q9uph7 homo sapien
23	297.5	18.2	805	13 Q91BA9	Q91ba9 xenopus lae
24	297	18.2	590	4 Q9BYR3	Q9byr3 homo sapien
25	286.5	17.5	925	11 Q9QZ00	Q9qz00 mus musculu
26	278.5	17.0	901	4 Q9H323	Q9h323 homo sapien
27	276.5	16.9	903	4 Q9BR81	Q9br81 homo sapien
28	266	16.3	765	11 Q70284	Q70284 mus musculu
29	265	16.2	657	11 Q35391	Q35391 mus musculu
30	257.5	15.7	849	5 Q15984	Q15984 bombyx mori
31	255	15.6	760	13 Q9PD06	Q9pdu6 xenopus lae
32	238	14.5	745	13 Q9BSJ5	Q9bsj5 brachydanto
33	169.5	10.4	719	5 Q963J8	Q963j8 caenorhabdl
34	169.5	10.4	823	5 Q45486	Q45486 caenorhabdl
35	169.5	10.4	825	5 Q9TYM0	Q9tym0 caenorhabdl
36	166	10.1	531	11 Q9Z3L9	Q9z3l9 rattus norv
37	158.5	9.7	575	13 Q91AU1	Q91au1 brachydanto
38	157.5	9.6	199	4 Q9Y4L8	Q9y4l8 homo sapien
39	157	9.6	103	6 Q9N110	Q9n110 ovis aries
40	156.5	9.6	534	11 Q9Z3L8	Q9z3l8 rattus norv
41	156	9.5	565	11 Q9Z4H3	Q9z4h3 rattus norv
42	154	9.4	87	13 Q90692	Q90692 gallus gall
43	153.5	9.4	852	5 Q9NC54	Q9nc54 mya arenari
44	149	9.1	108	11 Q90Z94	Q90z94 cavia porce
45	142	8.7	805	13 Q9BSN3	Q9bsn3 gallus gall

#### ALIGNMENTS

RESULT 1  
ID Q9QX54 PRELIMINARY; PRT; 630 AA.  
AC Q9QX54;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPXIXA-INDUCIBLE FACTOR 3 ALPHA (FRAGMENT).  
GN HIF3A.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gu Y.-Z., Moran S.M., Hogenesch J.B., Wartman L., Bradfield C.A.;  
RT "Cloning and Characterization of a Third Hypoxia Inducible Factor,  
RT HIF3-alpha.";  
RT J. Biol. Chem. 0:0-0(1999).  
RL EMBL; AF079153; AAF21782.1;  
DR EMBL; AF079140; AAF21782.1; JOINED.  
DR EMBL; AF079141; AAF21782.1; JOINED.  
DR EMBL; AF079142; AAF21782.1; JOINED.  
DR EMBL; AF079143; AAF21782.1; JOINED.  
DR EMBL; AF079144; AAF21782.1; JOINED.  
DR EMBL; AF079145; AAF21782.1; JOINED.  
DR EMBL; AF079146; AAF21782.1; JOINED.  
DR EMBL; AF079147; AAF21782.1; JOINED.  
DR EMBL; AF079148; AAF21782.1; JOINED.  
DR EMBL; AF079149; AAF21782.1; JOINED.  
DR EMBL; AF079150; AAF21782.1; JOINED.  
DR EMBL; AF079151; AAF21782.1; JOINED.  
DR EMBL; AF079152; AAF21782.1; JOINED.  
DR MGI:1859778; Hif3a.  
DR InterPro: IPR001092; HLH\_dim.  
DR InterPro: IPR003015; HLH\_Myc.  
DR InterPro: IPR001067; Nucleinslocator.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.

DR Pfam: PF00989; PAS: 2.  
DR PRINTS: PRO0785; NCRRNSLOCATR.  
DR SMART: SM00353; HLH: 1.  
DR SMART: SM00086; PAC: 1.  
DR SMART: SM00091; PAS: 2.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
FT NON\_TER 630 630  
SEQUENCE 630 AA: 69624 MW: 828832CA8E6D45B6 CRC64:

query Match	52.3%;	Score 855.5;	DB 11;	Length 630;
Best Local Similarity	68.4%;	Pred. No. 1.3e-77;		
Matches 182;	Conservative	5;	Mismatches 24;	Indels 55;
				Gaps 4

```

OY      9  RSNLRLKREKRDAAARSRSOSETVLQALHTLPFAGVSAHLDKASIMLTITSYLMHR  68
Db      7  RSNLRLKREKRDAAARSRSOSETVLQALHTLPFAGVSAHLDKASIMLTITSYLMHR  66
OY      69  LCAAGKFKGRATGRLLRPGCGEGRNGHTHRGRGSLPYGCKQADBPQSVLDC-----  120
Db      67  LCAAGENN-----QVEKGSGPLDACYKALEGF  94
OY      121  -----SSLIHNPTPGTNGS-LDELIGSTFDFTHPDQDELOALPRPNLSKKKL  170
Db      95  VAVLTAGSDMAVYLSBNYSKHLGLSLELEIGSTFTDTHPCQDELOALPRPNLSKKKL  154
OY      171  EAPTRHPSLSMRKSTLYSRGRFTLLMKATKTVLHCSGIMRAYKRPAPTSFAGSPRSRPEPL  230
Db      155  EAPTRHPSLSMRKSTLYSRGRFTLLMKATKTVLHCSGIMRAYKRPAPTSFAGSPRSRPEPL  214
OY      231  QCLVYLCEAIR-----QLPFHDGATL  251
Db      215  QCLVYLCEAIRPASLEPLGRGFL  240

```

RESULT	2			
Q992215		PRELIMINARY;	PRT;	662 AA.
AC	Q992215;			
DT	01-MAY-1999 (TREMBLrel, 10, Created)			
DT	01-MAY-1999 (TREMBLrel, 10, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel, 19, Last annotation update)			
DE	HYPOXIA INDUCIBLE FACTOR THREE ALPHA.			
CS	HIF3A.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	NCBI_Taxid=10090;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RA	Gu Y.Z., Moran S.M., Hogenesch J.B., Wartman L., Bradfield C.A.;			
RT	"Molecular characterization and chromosomal localization of a third			
RT	RT alpha-class hypoxia induced factor subunit, HIF3alpha.";			
RL	Gene Expr. 0:0-(1998).			
DR	EMBL: AF060194; AAC72734.1; -			
DR	MED: MGI:1859778; Hif3a.			
DR	InterPro: IPR001092; HLH dim.			
DR	InterPro: IPR003015; HLH_My.			
DR	InterPro: IPR001067; NucleinLocator.			
DR	InterPro: IPR001610; PAC.			
DR	InterPro: IPR000014; PAC.			
DR	Pfam: PF00989; PAS; 2.			
DR	PRINTS: PR00785; NCTRSLOCATR.			
DR	SMART: SM00353; HLH; 1.			
DR	SMART: SM00086; PAC; 2.			
DR	SMART: SM00091; PAS; 2.			
DR	PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.			
DR	SEQUENCE 662 AA; 73013 MW; 58740A1B69353B5 CRC64;			

Query Match	52.3%;	Score 855.5;	DB 11;	Length 662;
Best Local Similarity	68.4%;	Pred. No. 1.4e-77;		
Matches 182;	Conservative 5;	Mismatches 24;	Indels 55;	Gaps 4;

```

QY      9  RSNTELREKSRDAAARSRSOETEVIVOLATLPFARGVSAHLDKASIMULTISYLMHR  68
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      7  RSNTELREKSRDAAARSRSOETEVIVOLATLPFARGVSAHLDKASIMULTISYLMHR  66
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      69  LCAAGKGRGRATGRLLPEGPGGFRHGHRRGRHGLPVGCKQOAPGPOSDVLC-----  120
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      67  LCAAGEWN-----QVEKGEGIPDACYLKALEGF  94
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      121 -----SSSLIHNPPTGNTFS-TELIGHSTFDPLHPCDOEELQALPRPLSKKL  170
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      95  VMYLTAEQDMYILSEBNYSKHULGLSQLELTIGHSITDFHPCQDEELQALPRPLSKKL  154
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      171  EAPTERHFSLMKSKTILTSRGRTLNLKAAWVVLHCSGHMRAVYKPPAOTSPPAGSRSEPL  230
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      155  EAPTERHFSLMKSKTILTSRGRTLNLKAAWVVLHCSGHMRAVYKPPAOTSPPAGSRSEPL  214
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      231  QCLVYLICETAIIP-----QLPFDHGATL  251
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      215  QCLVYLICETAIIPASLEPPCLGRGAFL  240
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

	RESULT	3
Q96K34		
ID	Q96K34	PRELIMINARY; PRT; 632 AA.
AC	Q96K34;	
DT	01-DEC-2001 (TREMBLrel. 19, Created)	
PT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	CDNA FLJ14819 f1s, CLONE OVARC1000241, MODERATELY SIMILAR TO	
DE	HYPOMYXA-INDUCIBLE FACTOR 1 ALPHA.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi	
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=OVARIAN CARCINOMA:	
RA	Iisogaai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,	
RA	Nishikawa T., Nagai K., Sugeno S., Ishibashi T., Fujimori K.,	
RA	Saito H., Kimura M., Watanabe M., Hiraoxa S., Ishii S., Kawai Y.,	
RA	Tanai K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagaharai K.,	
RA	Masuhio Y., Kanehori K.;	
RT	"NEDO human cDNA sequencing project."	
RL	Submitted (MAY-2001) to the EMBL/genbank/DBJ databases.	
DQ	EMBL; AK027725; BAB55324.1, --	
DR	SEQUENCE 632 AA: 68934 MW; A19FIED3D05E7A71 CRC64;	

Query Match	50.8%;	Score 831;	DB 4;	Length 632;
Best Local Similarity	68.2%;	Pred. No. 3.8e-75;		
Matches 178;	Conservative 11;	Mismatches 30;	Indels 42;	Gaps 5

[illegible]



```

RESULT 4
ID Q9JHS2 PRELIMINARY; PRT: 662 AA.
AC Q9JHS2;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOXIA INDUCIBLE FACTOR 3 ALPHA.
GN HIF-3A.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=21134367; PubMed=11237857;
RA Kletzmann T., Cornesse Y., Brechtel K., Modaresi S., Jungermann K.;
RT "Perinatal expression of the mRNA of the three hypoxia-inducible
RT factor a-subunits HIF-1a, HIF2a and HIF3a in rat liver.";
RL Biochem. J. 354:531-537(2001).
DR EMBL: AJ277827; CAB96611.1; -
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001067; Nucleinslocator.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR PRINTS: PR00785; NCTRNLOCATR.
DR SMART: SM00353; HLH.1.
DR SMART: SM00086; PAC.1.
DR SMART: SM00091; PAS.2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ SEQUENCE 662 AA; 72887 MW; AC9672E340544010 CRC64;

```

```

Query Match 50.2%; Score 820.5; DB 11; Length 662;
Best Local Similarity 68.1%; Pred. No. 4.6e-74;
Matches 177; Conservative 9; Mismatches 31; Indels 43; Gaps 5;

QY 9 RSNTELREKSRDARSRSOETEVLYOLATLPFARGVSHLDKASIMRLTISYLRMR 68
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 7 RSTTELREKSRDARSRSOETEVLYOLATLPFARGVSHLDKASIMRLTISYLRMR 66
QY 69 LCAAGKRGATGRLDPGPGFRHGTTHRGHGLPVGKCO-----QAPGPOSV 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 67 LCAAG-----ENQYKKEGEPDADCYLKALEGFVWVLTAEQDAY 106
QY 118 DLCSSTLIHNPPTGNTS-LELIGHSTFDTHPCDOELDALTPRPWLSSKKLEAPTER 176
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 107 -----LSENVSKHLGSQLLELGHSTFDTHPCDOELDALTPRSLSKKSEAATGR 160
QY 177 HFSLRMKSTLTSRGRITLTKATWKVLCSGHMRAYKPPAQTSPAGSPRSPPLQCLVLI 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 161 HFSLRMKSTLTSRGRITLTKATWKVLCSGHMRAYKPPAQTSPAGSPRSPPLQCLVLI 220
QY 237 CEALP-----QLPHDQATL 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 221 CEALPRAPELPPGAGFL 240

RESULT 5
Q9Y2N7 PRELIMINARY; PRT: 667 AA.
ID Q9Y2N7;
AC Q9Y2N7;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE HOMOLOG OF HYPOXIA INDUCIBLE FACTOR THREE ALPHA
DE (HYPOXIA-INDUCIBLE FACTOR-3 ALPHA).
GN HIF-3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K.J., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Barnes J.,
RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Altix C., Andreise T., Frankheim M., Amico-Keller G., Coeffield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Ariellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 1.9 Mb region in 19q13.2 between APOE and
RT D19S412.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX Hara S., Kobayashi C., Hamada J., Imura N.;
RT "Characterization of human hypoxia-inducible factor-3 alpha.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007193; AAD22668.1; -
DR EMBL: AB054067; BAB9689.1; -
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001067; Nucleinslocator.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00989; PAS.2.
DR PRINTS: PR00785; NCTRNLOCATR.
DR SMART: SM00353; HLH.1.
DR SMART: SM00086; PAC.1.
DR SMART: SM00091; PAS.2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ SEQUENCE 667 AA; 72404 MW; 67B8794FEF9DCCFAB CRC64;

```

```

Query Match 49.0%; Score 802; DB 4; Length 667;
Best Local Similarity 68.0%; Pred. No. 3.4e-72;
Matches 172; Conservative 11; Mismatches 28; Indels 42; Gaps 5;

QY 9 RSNTELREKSRDARSRSOETEVLYOLATLPFARGVSHLDKASIMRLTISYLRMR 68
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 7 RSTTELREKSRDARSRSOETEVLYOLATLPFARGVSHLDKASIMRLTISYLRMR 66
QY 69 LCAAGKRGATGRLDPGPGFRHGTTHRGHGLPVGKCO-----QAPGPOSV 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 67 LCAAG-----ENQYKKEGEPDADCYLKALEGFVWVLTAEQDAY 106
QY 118 DLCSSTLIHNPPTGNTS-LELIGHSTFDTHPCDOELDALTPRPWLSSKKLEAPTER 176
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 107 -----LSENVSKHLGSQLLELGHSTFDTHPCDOELDALTPROTLSSRKVEAPTER 160
QY 177 HFSLRMKSTLTSRGRITLTKATWKVLCSGHMRAYKPPAQTSPAGSPRSPPLQCLVLI 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 161 HFSLRMKSTLTSRGRITLTKATWKVLCSGHMRAYKPPAQTSPAGSPRSPPLQCLVLI 220
QY 237 CEALPPLPFHDA 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 221 CEALP-----HPGS 229

RESULT 6
Q9HA12 PRELIMINARY; PRT: 648 AA.
ID Q9HA12;
AC Q9HA12;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE HYPOPHETICAL 70.0 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EMBryo;  
RA Itogeti T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiraori A., Sudo H.,  
RA Magatsuna M., Hosori T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,  
RA Niinomiya K., Iwayanagi T.;  
RT "NEO human cDNA sequencing project";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK021653; BAB13865.1; -;  
DR InterPro; IPR001610; PAC.  
DR InterPro; IPR000014; PAS.  
DR Pfam; PF00989; PAS; 2.  
DR SMART; SM00086; PAC; 1.  
DR SMART; SM00091; PAS; 2.  
SQ SEQUENCE 648 AA; 6994 MW; EBERC744BC3F148E CRC64;

Query Match 35.0%; Score 573; DB 4; Length 648;  
Best Local Similarity 61.0%; Pred. No. 4; ie-49;  
Matches 125; Conservative 11; Mismatches 27; Indels 42; Gaps 5;  
QY 57 MRLTSLYRMRHRCAGGKGRATGRLPRGPGFRHGRHGRHGLPVGKQ----- 109  
DB 1 MRLTSLYRMRHRCAGG-----EMNQVAGSGEPDLACVLEKALEGF 40  
QY 110 ----QAPGPOVDLCSSLIHNPTPGTNEF-LELIGHSIDFTHPCQOELODALTPRPN 164  
DB 41 VVNLTEGDVAY-----LSENVSKHLGSLQLEIGHSIDFTHPCQOELODALTPROQT 94  
QY 165 LSKKKLEAPTERHPSLRKSTLTSRGRTLNKATWVHLHSGHMRAYKPPAOTSPAGSP 224  
DB 95 LSRKKEAPTERHPSLRKSTLTSRGRTLNKATWVHLHSGHMRAYKPPAOTSPAGSP 154  
QY 225 RSEPPLOCLVLCGAIPLPFDHGA 249  
DB 155 DSEPPLOCLVLCGAIPLPFDHGA 175

RESULT 7  
Q9HAM5 PRELIMINARY; PRT; 199 AA.  
AC Q9HAM5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE HYPOTHETICAL 21.6 KDA PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EMBryo;  
RA Itogeti T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fuji A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara K., Takeuchi K.,  
RA Arita M., Nabejima T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;  
RT "NEO human cDNA sequencing project";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK021421; BAB13819.1; -;  
SQ SEQUENCE 199 AA; 21597 MW; 646CB066A09B47C3 CRC64;

Query Match 31.7%; Score 518.5; DB 4; Length 199;  
Best Local Similarity 68.0%; Pred. No. 2; ie-44;  
Matches 104; Conservative 9; Mismatches 13; Indels 27; Gaps 2;

QY 120 CSSSLIHNPTPGTNEF-----STELIGHSTFDFTHPCQOELO 156  
DB 14 CTSMLTRCPSPAPASAPTRPLSCASPSACTASAPOLDELGHSTFDFTHPCQOELO 73  
QY 157 DALTPRPNLSSKKLEAPTERHPSLRKSTLTSRGRTLNKATWVHLHSGHMRAYKPPA 216  
DB 74 DALTPROTLSSRKYVEAPTEFCFSLRKSTLTSRGRTLNKATWVHLHSGHMRAYKPPA 133  
QY 217 QTSPPAGSPSEPPLOCLVLCGAIPLPFDHGA 249  
DB 134 QTSPPAGSPSEPPLOCLVLCGAIPLPFDHGA 162

RESULT 8  
Q9YIB9 PRELIMINARY; PRT; 811 AA.  
AC Q9YIB9;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL INDUCIBLE FACTOR-1 ALPHA.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HEART;  
RA Takahashi T.;  
RT "Molecular cloning and expression of an avian cDNA for hypoxia-inducible factor-1 alpha in embryonic ventricular myocytes";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB013746; BAA34234.2; -;  
DR InterPro; IPR001092; HLH\_dim.  
DR InterPro; IPR001610; PAC.  
DR InterPro; IPR000014; PAS.  
DR Pfam; PF00785; PAC; 1.  
DR SMART; SM00353; HLH; 1.  
DR SMART; SM00086; PAC; 1.  
DR SMART; SM00091; PAS; 2.  
SQ SEQUENCE 811 AA; 90542 MW; D14CD9FC98F064CB CRC64;

Query Match 30.7%; Score 503; DB 13; Length 811;  
Best Local Similarity 49.8%; Pred. No. 6; ie-42;  
Matches 119; Conservative 20; Mismatches 72; Indels 28; Gaps 5;

QY 9 RSNTELRKESRDPAASRSQETVELYQLAHTLPFANGVAHLDKASIMRLTSLYLRMR 68  
DB 12 RISSEKREKSRDPAARCRSESEVEFEELAHQPLPRTVSAHLDKASIMRLTSLYLRMR 71  
QY 69 LCAAGCKRGAT-----GRLPRGPGGFRHGRHGLPVKCGQAPGQSVDCSS 123  
DB 72 LLDGGELETANNEKELNCYLKAALDGFVAVLSEDC-----DMLYMS 113  
QY 124 LIHNPTPG-TNFSLELIGHSTFDFTHPCQOELODALTPRPNLSSKKLEAPTERHPSLRM 182  
DB 114 ENNVKCKMGLQF--DLGHGVDFPTHCHDEBLREMLTHNGVKKGKEQNTERSFPLRM 171  
QY 183 KSLTSGRGLNLKATWVHLHSGHMRAYKPPAOTSPAGSPSEPPLOCLVLCGAIPL 241  
DB 172 KCLTSGRGLNLKATWVHLHSGHMRAYKPPAOTSPAGSPSEPPLOCLVLCGAIPL 228

RESULT 9  
Q9XTA5 PRELIMINARY; PRT; 823 AA.  
AC Q9XTA5;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL INDUCIBLE FACTOR-1 ALPHA.

GN HIF-1 ALPHA.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=ARTERY;  
 RX MEDLINE=99255430; PubMed=10320777;  
 RA Hara S., Kobayashi C., Imura N.;  
 RT "Molecular cloning of cDNAs encoding hypoxia-inducible factor (HIF)-  
 RL 1alpha and -2alpha of bovine arterial endothelial cells.";  
 DR Biochim. Biophys. Acta 1445:237-243(1999).  
 DR EMBL; AB018398; BAA78675.1; -.  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001321; HypoxIndf1A.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam; PF00785; PAS; 1.  
 DR Pfam; PF00989; PAS; 2.  
 DR PRINTS; PRO1080; HYPOXIAIF1A.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
 SQ SEQUENCE 823 AA; 92127 MW; 12674E467A61B1A1 CRC64;

Query Match Best Local Similarity 30.5%; Score 499; DB 6; Length 823;  
 Matches 113; Conservative 26; Mismatches 58; Indels 48; Gaps 6;

QY 11 NTELRRKESRDAAARSRSQETEVLYQLAHTLPFAKGVSAHLDKASIMRLTISYLRMRHC 70  
 :::::::::::::::::::::  
 DB 14 SSERKRKSRDAAARSRSKSESEVFYELAHQPLPHNVSSHLDKASVRLTISYLRKLL 73  
 :::::::::::::::::::::  
 QY 71 AAGKRGATGRLLPBGEGGFRHGRHGRGLPVGKCOA-----PGPSVDLC 120  
 :::::::::::::::::::::  
 DB 74 DAG-----DLDEDMKAKMNCFYLKALDGFVAVLTD 105  
 :::::::::::::::::::::  
 QY 121 SSSLINPTPG-TNFSLELIGHISIFDFIHPDOEELDALTPRPNLKKLEAPTR 174  
 :::::::::::::::::::::  
 DB 106 DGDATVYSDNNKTKMGLTF--ELTGHVSFDFTHPCDHEKREMLTHRNGLYKKKEONT 163  
 :::::::::::::::::::::  
 QY 175 ERHPSLRKSTLTSRGRTLNKAATWVLCSGHMRAYKPPAOTSPAGSPRSEPPLOCLIV 234  
 :::::::::::::::::::::  
 DB 164 QRSFELRMKCTLTSGRTMINKSATWVLCGHITGHIVDTNSNQSGC--YKKPPMTCLV 221  
 :::::::::::::::::::::  
 QY 235 LICEAIP 241  
 :::::::::::::::::::::  
 DB 222 LICEPIP 228

RESULT 10  
 035800 PRELIMINARY; PRT; 825 AA.  
 AC 035800;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOXIA-INDUCIBLE FACTOR 1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WISTAR;  
 RX MEDLINE=21134367; PubMed=11237857;  
 RA Kietzmann T., Cornesse Y., Brechtel K., Modaresi S., Jungermann K.;  
 RT "Pervious expression of the mRNA of the three hypoxia-inducible

RT factor a-subunits HIF-1a, HIF2a and HIF3a in rat liver.";  
 RL Biochem. J. 354:531-537(2001).  
 DR EMBL; Y09507; CAA70701.1; -.  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001321; HypoxIndf1A.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam; PF00785; PAS; 1.  
 DR Pfam; PF00989; PAS; 2.  
 DR PRINTS; PRO1080; HYPOXIAIF1A.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
 SQ SEQUENCE 825 AA; 92319 MW; C4109A57E38667E9 CRC64;

Query Match Best Local Similarity 30.4%; Score 497; DB 11; Length 825;  
 Matches 113; Conservative 29; Mismatches 72; Indels 28; Gaps 5;

QY 6 QVRNTELRKESRDAAARSRSQETEVLYQLAHTLPFAKGVSAHLDKASIMRLTISYLR 65  
 :::::::::::::::::::::  
 DB 9 EKKMSSEKRSRDAAARSRSKSESEVFYELAHQPLPHNVSSHLDKASVRLTISYLR 68  
 :::::::::::::::::::::  
 QY 66 MRLCAAC-----GKRGATGRLLPBGEGGFRHGRHGRGLPVGKCOAAGPSVDLC 120  
 :::::::::::::::::::::  
 DB 69 VRKLDAGDLDIEDMKAKMNCFYLKAPDGFVAVLTDG-----DRI 110  
 :::::::::::::::::::::  
 QY 121 SSSLINPTPG-TNFSLELIGHISIFDFIHPDOEELDALTPRPNLKKLEAPTRHRS 179  
 :::::::::::::::::::::  
 DB 111 YISDNNKTKMGLTF--ELTGHVSFDFTHPCDHEKREMLTHRNGLYKKKEONT 168  
 :::::::::::::::::::::  
 QY 180 LRKSTLTSRGRTLNKAATWVLCSGHMRAYKPPAOTSPAGSPRSEPPLOCLIVLCEA 239  
 :::::::::::::::::::::  
 DB 169 LRKCTLTSGRTMINKSATWVLCGHITGHIVDTNSNQSGC--YKKPPMTCLV 226  
 :::::::::::::::::::::  
 QY 240 IP 241  
 :::::::::::::::::::::  
 DB 227 IP 228

RESULT 11  
 09CYA8 PRELIMINARY; PRT; 258 AA.  
 AC 09CYA8;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOXIA INDUCIBLE FACTOR 1, ALPHA SUBUNIT.  
 GN HIF1A.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arkawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fetschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Botelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guenlinich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombearts P.,

Query Match	30.3%;	Score 495;	DB 4;	Length 735;
Best Local Similarity	46.2%;	Pred. No. 3.5e-41;		
Matches 114;	Conservative 26;	Mismatches 59;	Indels 48;	Gaps 6

Query Match	30.3%;	Score 495;	DB 4;	Length 826;
Best Local Similarity	46.2%;	Pred. No. 4e-41;		
Matches 114;	Conservative 26;	Mismatches 59;	Indels 48;	Gaps 6;

QY	11	NTLARKESRDAARSRROETEVYLQOLAHPTLPFAGVSAHLDKASIMRLTISYLRNRHC	70
QY	11	NTLARKESRDAARSRROETEVYLQOLAHPTLPFAGVSAHLDKASIMRLTISYLRNRHC <td>70</td>	70
Db	14	SSERRKESRDAARSRRRKESSEVEFYELAHQPLPRPNVSSHLDKASVRLTISYLRKLL	73
QY	71	AAGGKRGATGRLPEGGGFGFRHGTNRGRHCLPFGKCOQA-----PGQSYDLC	120
Db	74	DAG-----DLDIEDDKKAQMNCFYLKALDGFVWVLT	105
QY	121	SSSLIH-----NPPPG-TNFSLELIGHSIFDEIHPDCOELODALTPRPNLSKKKLEAPT	174
Db	106	DGDMIYISDNVNYKMGTLQF--ELTGHVSFDEPTPHCDHEMKREMLTHRNGLVKKGQEW	163
QY	175	ERHFSLRKSKTITSNGRFLNKAATWKVLCGNGHAKRKPRAQTSAPGSPREPPLOCLV	234
Db	164	QSFELRKKCVLTISGRGTRMINKSATWKVLCGTHIHYDTNSNPQCG--YKRPPTCLV	221
QY	235	LICEAIP 241	
Db	222	LICEPIP 228	
RESULT 14			
ID	Q9WTU9	PRELIMINARY:	PRT: 823 AA.
AC	Q9WTU9:		
DT	01-NOV-1999 (TREMBLrel. 12, Created)		
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	HYOXA-INDUCIBLE FACTOR-1 ALPHA.		
GN	HIF1A.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SPRAGUE DAWLEY; TISSUE=KIDNEY;		
RX	MEDLINE=21417706; PubMed=11526200;		
RA	Zou A.P., Yang Z.Z., Li P.L., Cowley AW J.R.;		
RT	"Oxygen-dependent expression of hypoxia-inducible factor-1alpha in		
RL	renal medullary cells of rats.";		
DR	Physiol. Genomics (Online) 6:159-168(2001).		
DR	EMBL; AF057308; AAD24413.1;--		
DR	InterPro: IPR001092; HLH_dlm.		
DR	InterPro: IPR003015; HLH_MYC.		
DR	InterPro: IPR001321; Hypoxindf1A.		
DR	InterPro: IPR000014; PAS.		
DR	Pfam: PF00785; PAC; 1.		
DR	Pfam: PF00989; PAS; 2.		
DR	PRINTS; PRO1080; HYPOXIA1F1A.		
DR	SMART; SM00353; HLH; 1.		
DR	SMART; SM00086; PAC; 1.		
DR	SMART; SM00091; PAS; 2.		
DR	PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN.1.		
SO	SEQUENCE 823 AA; 92116 MW; ODC907B530766FIF CRC64;		
Query Match 30.0%; Score 491; DB 11; Length 823;			
Best Local Similarity 45.4%; Pred. No. 1e-40; Mismatches 61; Indels 48; Gaps 6			
QY	9	RSNTELRKESRDAARSRROETEVYLQOLAHPTLPFAGVSAHLDKASIMRLTISYLRMR	68
Db	13	RMSRRKESRDAARSRRRKESSEVEFYELAHQPLPRPNVSSHLDKASVRLTISYLRVAK	72
QY	69	LCAAGGKRGATGRLPEGGGFGFRHGTNRGRHGLPVGKCOQA-----PGQSYD	118
Db	73	LLGAG-----DLDIEDDKKAQMNCFYLKALDGFVWVLT	104
QY	119	LCSSSLIH-----NPPPG-TNFSLELIGHSIFDEIHPDCOELODALTPRPNLSKKKLEA	172
Db	105	TDDGMITYISDNVNYKMGTLQF--ELTGHVSFDEPTPHCDHEMKREMLTHRNGPVVRKQEO	162

```

OY      173 PTERFSLRMSTLTSRRTLNKATTKVLHSCGHMAKYPRPQTSAGSPRESEPPLOC 232
        ||| | | | | | | | | | | | | | | | | : | | | | |
Db      163 NORSFELRMCTLTLSRRTNMRKSATWKVLCHTGHIHVDTSSNQPOCG--YKKPWTC 220
OY      233 LVLICETAP 241
        ||| | | | |
Db      221 LVLICETIP 229

RESULT  15
O9JHS1  ID O9JHS1 PRELIMINARY; PRF; 874 AA.
AC       Q9JHS1:
DT       01-OCT-2000 (TREMBLrel. 15, Created)
DT       01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE       01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE       HYPOXIA INDUCIBLE FACTOR 2 ALPHA.
GN       HIF-2A.
OS       Rattus norvegicus (Rat).
OC       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC       Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX       NCBI_TaxID=10116;
RX       [1]
RC       SEQUENCE FROM N.A.
RC       TISSUE=LIVER;
RX       MEDLINE=21134367; PubMed=11237857;
RT       Kietzmann T., Cornesse Y., Brechtel K., Modarressi S., Jungermann K.;
RT       "Particulous expression of the mRNA of the three hypoxia-inducible
RL       factor a-subunits HIF-1a, HIF2a and HIF3a in rat liver.";
RL       Biochem. J. 354:531-537(2001).
DR       EMBL: AJ277828; CAB96612.1; -
DR       InterPro: IPR001092; HLH.dlm.
DR       InterPro: IPR003015; HLH_Myc.
DR       InterPro: IPR001067; NucleinsLocator.
DR       InterPro: IPR001610; PAC.
DR       InterPro: IPR000014; PAS.
DR       Pfam: PF00785; PAC; 1.
DR       Pfam: PF00989; PAS; 2.
DR       PRINTS: PRO0785; NCTRNSLOCATR.
DR       SMART: SM00353; HLH; 1.
DR       SMART: SM00086; PAC; 1.
DR       SMART: SM00091; PAS; 2.
DR       PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SO       SEQUENCE 874 AA; 96718 MW; A1FF08EB24369796 CRC64;
```

[illegible]

Search completed: September 8, 2002, 07:41:25  
Job time: 443 sec

---